

Journal of
Geophysical
Research

VOLUME 66 NOVEMBER 1961 NUMBER 11

PUBLISHED BY
THE AMERICAN GEOPHYSICAL UNION

Journal of Geophysical Research

An International Scientific Publication

OFFICERS OF THE UNION

THOMAS F. MALONE, *President*
GEORGE P. WOOLLARD, *Vice President*
A. NELSON SAYRE, *General Secretary*
WALDO E. SMITH, *Executive Secretary*

OFFICERS OF THE SECTIONS

Geodesy

FLOYD W. HOUGH, *President*
CHARLES A. WHITTEN, *Vice President*
BUTFORD K. MEADE, *Secretary*

Seismology

JAMES A. PEOPLES, JR., *President*
JACK E. OLIVER, *Vice President*
BENJAMIN F. HOWELL, JR., *Secretary*

Meteorology

MORRIS NEIBURGER, *President*
HENRY G. HOUGHTON, *Vice President*
WOODROW C. JACOBS, *Secretary*

Geomagnetism and Aeronomy

C. T. ELVEY, *President*
E. H. VESTINE, *Vice President*
J. HUGH NELSON, *Secretary*

Oceanography

DONALD W. PRITCHARD, *President*
ROBERT S. ARTHUR, *Vice President*
ARTHUR E. MAXWELL, *Secretary*

Volcanology, Geochemistry, and Petrology

HATTEN S. YODER, JR., *President*
EDWARD D. GOLDBERG, *Vice President*
DAVID R. WONES, *Secretary*

Hydrology

WILLIAM C. ACKERMANN, *President*
DAVID K. TODD, *Vice President*
RALPH N. WILSON, *Secretary*

Tectonophysics

LOUIS B. SLICHTER, *President*
DAVID T. GRIGGS, *Vice President*
IRIS BORG, *Secretary*

BOARD OF EDITORS

Editors: PHILIP H. ABELSON and J. A. PEOPLES, JR.

ASSOCIATE EDITORS

1959-1961

HENRI BADER	T. NAGATA
K. E. BULLEN	FRANK PRESS
CONRAD P. MOOK	A. NELSON SAYRE
WALTER H. MUNK	MERLE A. TUVE
JAMES A. VAN ALLEN	

1960-1962

JULIUS BARTELS	L. A. MANNING
V. V. BELOUSSOV	TOR J. NORDENSON
E. G. BOWEN	E. N. PARKER
JOHN E. CHAPPELEAR	GEORGE P. RIGSBY
G. D. GARLAND	WALTER O. ROBERTS
GORDON J. F. MACDONALD	C. N. TOUART

JAMES R. WAIT

1961-1963

FRANKLIN I. BADGLEY	ROBERT O. REID
HENRY G. BOOKER	BRUNO ROSSI
JOSEPH W. CHAMBERLAIN	GEORGE H. SUTTON
HERBERT FRIEDMAN	DAVID K. TODD
MARK F. MEIER	VICTOR VACQUIER

ARTHUR H. WAYNICK

The Editors of the *Journal of Geophysical Research* welcome original scientific contributions on the physics of the earth and its environment.

Manuscripts should be submitted in triplicate to J. A. Peoples, Jr., Department of Geology, University of Kansas, Lawrence, Kansas. Authors' institutions if in the United States or Canada, are requested to pay a publication charge of \$25 per page, which, if honored, entitles them to 100 free reprints.

Subscriptions to the *Journal of Geophysical Research* and *Transactions, AGU*, are included in membership dues.

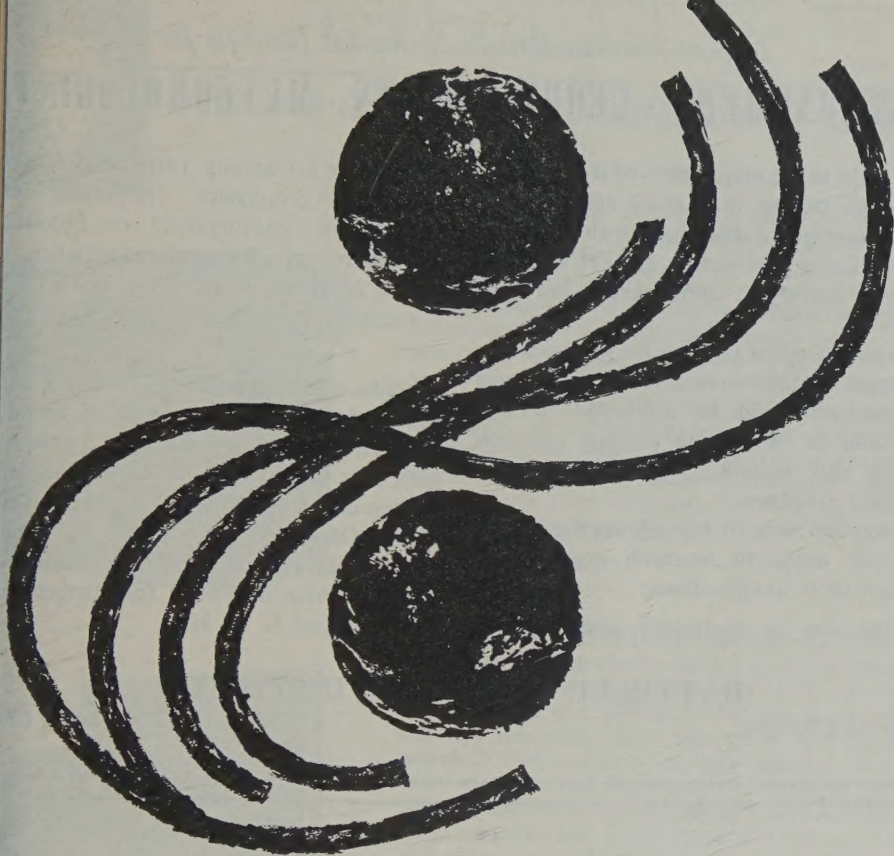
Nonmember subscriptions, *Journal of Geophysical Research*...\$30 for back volume of 1959, \$42 for back volume of 1960, \$5 for this issue; \$20 for the calendar year 1961.

Nonmember subscriptions, *Transactions, AGU*...\$4 per calendar year, \$1.25 per copy.

Subscriptions, renewals, and orders for back numbers should be addressed to American Geophysical Union, 1515 Massachusetts Ave., Northwest, Washington, D. C. Suggestions to authors are available on request. Advertising Representative: Howland and Howland Inc., 230 Park Ave., New York 17, N. Y.

Since January 1959 (Vol. 64, No. 1) the *Journal of Geophysical Research* has been published monthly by the American Geophysical Union, the U. S. National Committee of the International Union of Geodesy and Geophysics, organized under the National Academy of Sciences-National Research Council as the U. S. national adhering body. Publication of this journal is supported by the National Science Foundation and the Carnegie Institution of Washington. The new monthly combines the type of scientific material formerly published in the bi-monthly *Transactions, American Geophysical Union*, and the quarterly *Journal of Geophysical Research*. The *Transactions, American Geophysical Union*, continues as a quarterly publication for Union business and items of interest to members of the Union.

Published monthly by the American Geophysical Union from 1407 Sherwood Avenue, Richmond, Virginia. Second class postage paid at Richmond, Virginia.



Johann Kepler: *"The planets move in elliptical orbits about the sun, and the square of their period of revolution are proportional to the cube of their mean distances from the sun."*

interplanetary voyages fast becoming a reality, complete information regarding the velocity requirements for travel between planets is of vital importance. With these data available, it is possible to analyze propulsion requirements, plan system configurations, and conduct feasibility studies for any particular mission.

Lockheed Missiles & Space Company scientists have actually evolved a rapid-calculation method, utilizing a high-speed computer. This has produced literally thousands of orbits, velocity requirements, and elapsed time, for design studies of missions to and from both Mars and Venus—every tenth day from now until January, 1970.

As simple to analyze are many factors which make Lockheed Missiles & Space Company a wonderful place to live and work. Located in Sunnyvale and Palo Alto, California, on the beautiful San Francisco Peninsula, Lockheed is Systems Division for such programs as the DISCOVERER and MIDAS satellites and the POLARIS FBM. These, together with research and development projects in all disciplines, make possible a wide diversity of positions for creative engineers and scientists in chosen fields.

Do not investigate future possibilities at Lockheed? Write Research and Development Staff, Dept. M-14A, 962 West Evelyn Avenue, Sunnyvale, Calif. U.S. citizenship or existing Department of Defense industrial security clearance required. Equal Opportunity Employer.

LOCKHEED MISSILES & SPACE COMPANY

A GROUP DIVISION OF LOCKHEED AIRCRAFT CORPORATION

Program Manager for the Navy POLARIS FBM and the Air Force AGENA Satellite in the DISCOVERER and MIDAS programs. Other current programs include SAINT, ADVENT and such NASA projects as OGO, OAO, ECHO, and NIMBUS.

Offices: PALO ALTO, VAN NUYS, SANTA CRUZ, SANTA MARIA, CALIFORNIA • CAPE CANAVERAL, FLORIDA • HAWAII

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

Urgent Need by Battelle Memorial Institute for
ASTRONOMERS - GEOPHYSICISTS - METEOROLOGISTS

To assist in the preparation of state-of-the-art-type studies in various astronomical and geophysical disciplines with emphasis on defining the earth's natural environment—terrestrial, atmospheric and spatial.

In the process of reviewing and evaluating current world-wide research in these areas applicant should be interested in and capable of identifying existing research voids and designing appropriate experimental programs.

If you are interested in joining the largest private nonprofit research organization check these qualifications:

Bachelor or preferably advanced de-

grees in astronomy (astrophysics emphasis) geophysics (excluding exploration geophysics) or physical meteorology. Research experience in one of these areas, while not a prerequisite would be desirable.

Battelle offers excellent laboratory and library facilities, on-the-premises contact with virtually all sciences and technologies, plus fine living in a cultured metropolitan center. Battelle is located adjacent to The Ohio State University. Good salary and unusually complete benefits. Why not write for further details. Address L. G. Hill

BATTELLE MEMORIAL INSTITUTE

505 King Avenue

Columbus 1, Ohio

BULLETIN (IZVESTIYA), ACADEMY OF SCIENCES, U.S.S.R.
GEOPHYSICS SERIES

Subscriptions for 1961 series now available

This monthly Russian publication, perhaps the leading journal of Geophysics of the U.S.S.R., is being translated and published in an English edition for the year 1961 by the American Geophysical Union. The twelve numbers in Russian cover about 2000 pages. Published with the aid of a grant from the National Science Foundation.

Send subscriptions now to

AMERICAN GEOPHYSICAL UNION
1515 Massachusetts Avenue, N.W.
Washington 5, D. C., U.S.A.

Subscription rates: \$25.00 for the volume of 12 numbers (\$20.00 to individual members of AGU subscribing for personal use)
Numbers will be mailed as issued.

The English edition of this publication for 1957 has been translated and published for the American Geophysical Union by Pergamon Press. This volume may be ordered through the American Geophysical Union at a price of \$25.00. The 1958, 1959, and 1960 series are available at a price of \$25.00 for each volume of 12 numbers. Titles and authors of the papers contained in the series have been published in recent issues of the *Transactions, AGU*.

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

GEOTECH

PRECISION INSTRUMENTS for SEISMOLOGY—GEOPHYSICS

HELICORDER

An immediate data display, and up to 72 hours of data recorded on one 12" x 36" record are featured by this precision recorder. The penmotor(s) translates as the drum revolves, recording data on heat sensitive paper in helical fashion. Significant data is easily identified with one glance at the entire record. No dark room or delay for photo processing is necessary, and there is no ink to clog. One- and three-trace models are available. Drum speed is adjustable, and 8 to 72 hours of analog data in a band pass of 0 to 24 cps can be recorded on one permanent record. The companion amplifier shown is also available. The recorder can be rack-mounted or furnished with a cabinet for table mounting. Dependability is proven through years of trouble-free operation under extreme field conditions.

DEVELOCORDER

This is a combination 16 mm film recorder, automatic film processor, and film viewer. With 1 to 16 miniature Geotech galvanometers installed, Model 4000 will record up to 120 hours of data on a 200 ft. roll of film. Band pass is 0 to 120 cps depending upon film speed and galvanometers selected. With a cathode ray tube installed, Model 4000A will also record data across the film, as well as intensity-modulated spectrograms, Lissajous patterns, and other presentations. A film speed of 3 cm/min is standard, and speeds from 0.6 to 20 cm/min are available on special order. Self-contained chemicals process the film internally. Data is displayed magnified 10X on the 6" x 17" view-screen within 2 to 20 minutes after recording. A 2-directional, variable-speed drive is provided for viewing. Date and time are printed on the film at regular intervals.

PORTABLE STATION TIMING SYSTEM

This is a complete station timing system for precision recording and timing applications which has been transistorized and packaged in miniature modules for dependable, portable field use. It provides accurate clock time and programmed time marks, and 15 va of 60-cps power output, all with a stability of at least 5 parts in 10^7 per week. Individual modules, which can be obtained separately, are: an oven- and crystal-controlled frequency standard, a frequency divider, a strobe unit for comparing and adjusting system time with external time, a clock and time-mark programmer, a speaker for radio time signals, and an amplifier providing the output power for recorder drives and other devices. Dimensions are 19"W x 10.5"H x 9"D. Weight is 25 pounds. Power required is 2 amps, 24 volts, DC. These systems have operated in the field for over a year.

For detailed specifications, price, or delivery, write to:

THE GEOTECHNICAL CORP.

HILOH ROAD

GARLAND, TEXAS

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

PORTABLE, ACCURATE, EASY TO OPERATE Sprengnether's Blast and Vibration Seismograph

Ideal for recording all types of vibrations caused by blasting, pile driving, heavy industrial machinery and other sources of strong motion vibrations.

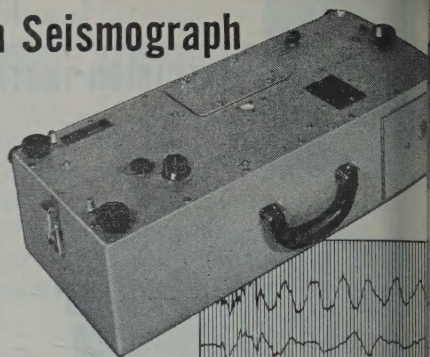
Portability (38 lbs. — 25 x 10 x 8 in.) Unit is self contained and free from external power source.

Extremely Accurate To guard against error, each instrument is tested and calibration data furnished. Frequency response, 3 to 200 cycles per second. Timing lines are across record at intervals of 0.02 seconds with accuracy of 0.1%.

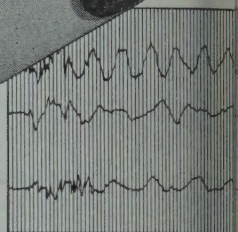
Easy to Operate: All controls are easily accessible. Instrument can be set up, leveled and made ready to operate within minutes.

Seismometer System: A mechanical, optical seismometer employing three independent pendulum systems with magnetic damping. System is contained within unit, hence, no need for external geophones.

Recording System: Photographic recording of all three components appearing on 2 $\frac{3}{4}$ inch wide paper. Cartridge type cameras are replaceable and can be pre-loaded to facilitate in the field camera replacement.



Write today for
complete
information.



OTHER SPECIFICATIONS

Natural Period (All Components)..... 0.75
Damping (Fraction of Critical)..... .55
Static Magnification..... *

*May be specified by purchaser from 50 to 200.
Two ranges in one instrument available.

Internationally Known Mfrs. of Seismological, Geophysical Instruments.

W.F. SPRENGNETHER INSTRUMENT CO., INC.
4567 SWAN AVENUE • ST. LOUIS 10, MO.

GEODESY AND CARTOGRAPHY, 1959 and 1960

(Geodezia i Kartografiya)

The leading monthly journal of Geodesy and Cartography in the USSR is being translated and published in an English edition, *for the years 1959 and 1960*, by the American Geophysical Union, aided by a grant from the National Science Foundation.

Subscription price, \$20.00 for each volume of 12 numbers

Send subscription requests to

AMERICAN GEOPHYSICAL UNION

1515 Massachusetts Avenue, N.W.

Washington 5, D. C., U.S.A.

Other publications of the AGU include

- JOURNAL OF GEOPHYSICAL RESEARCH (monthly), 4800 pages anticipated for 1961, subscription \$20.00
- TRANSACTIONS (quarterly), \$4.00 per calendar year
- GEOPHYSICAL MONOGRAPH SERIES (occasional volumes), No. 5 (issued in 1960) *Physics of Precipitation* (proceedings of the Cloud Physics Conference, Woods Hole, Massachusetts, June 3-5, 1959), 435 pp., \$12.50
- IZVESTIYA, Academy of Sciences, USSR, Geophysics series, English edition (monthly), \$25.00 per calendar year, available for 1957, 1958, 1959, 1960, and 1961.

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

IONOSPHERIC PROPAGATION and HF COMMUNICATIONS



If you have the background, the imagination and the desire to contribute to important programs in these fields, you are invited to join a carefully selected team of outstanding scientists and engineers now contributing significantly to current knowledge through advanced research.

These programs are being conducted at our ELECTRO-PHYSICS LABORATORIES in the SUBURBAN WASHINGTON, D. C. area, ideally located from the viewpoint of advanced study which may be conducted at one of several nearby universities; for readily available housing in pleasant residential neighborhoods; and for the general amenities of being offered by this important Metropolitan center. All qualified applicants will receive consideration for employment without regard to race, creed, color or national origin.

PRESENT NEEDS ARE FOR:

SENIOR IONOSPHERIC

PHYSICISTS Ph.D. preferred, with several years' experience in the study of Ionospheric phenomena. Should be familiar with present knowledge of upper atmosphere physics and possess an understanding of current programs using rockets and satellites for studies in F-region and beyond. Qualified individuals with supervisory abilities will have an exceptional opportunity to assume project leadership duties on HF projects already under way involving F-layer propagation studies, backed by a substantial experimental program.

SENIOR DEVELOPMENT

PHYSICISTS Advanced degree in Physics or E.E. preferred. Must be familiar with latest techniques in the design of advanced HF receivers and transmitters and possess working knowledge of modern HF networks employing waveguides and metallic tape cores. Strong theoretical background and modern linear circuit theory desired. Will carry out laboratory development and implementation of new HF communications systems.

SENIOR ELECTRONIC

ENGINEERS Advanced degree in E.E. preferred. Must be familiar with conventional pulse circuit designs and applications. Technical background should include substantial experience in data process and data recovery systems using

both analog and digital techniques. Knowledge of principles and application of modern information theory including correlation techniques helpful. Will be responsible for the design of sub-systems.

JUNIOR ELECTRONIC

ENGINEERS To assist Senior Engineers and Scientists in the development of HF communications and data process equipment. Should have formal electronics schooling and 2 years' experience in circuit design checkout or analysis of HF communications, Radar Pulse, Analog/Digital or Data Recovery equipment. Construction of prototypes of new and interesting equipment and design of individual components of communications and data processing systems will comprise the major efforts of selected applicants.

FIELD STATION

ENGINEERS B.S.E.E. or equivalent, consisting of combined civilian or military technical school, with work experience. Presently employed as a field engineer or project engineer with a valid 1st or 2nd Class FCC license and a good command of some of the following: Radar, preferably high power; HF long-distance communications systems; Tropospheric or Ionospheric scatter systems. Must be willing to accept assignments in areas where dependents are not permitted for periods of up to one year. Differential paid for overseas assignments.

For a prompt reply to your inquiry, please forward resume in confidence to: Dept. C-1, W. T. WHELAN, Director of Research & Development

ACF ELECTRONICS
DIVISION

ACF INDUSTRIES

HYATTSVILLE, MARYLAND

Journal of Geophysical Research

BACK ISSUES AVAILABLE

Volume 64, 1959 **Total 2488 pages**

Complete Volume	\$30.00
January, 132 pp.	\$2.00
February, 138 pp.	\$2.00
March, 112 pp.	\$2.00
April, 106 pp.	\$2.00
May, 98 pp.	\$2.00
June, 110 pp.	\$2.00
July, 168 pp.	\$2.00
August, 268 pp.	\$4.00
September, 230 pp.	\$3.00
October, 284 pp.	\$4.00
November, 390 pp.	\$5.00
December, 452 pp.	\$6.00

Volume 65, 1960 **Total 4248 pages**

Complete Volume	\$42.00
January, 384 pp.	\$5.00
February, 414 pp.	\$5.00
March, 284 pp.	\$4.00
April, 248 pp.	\$4.00
May, 314 pp.	\$4.00
June, 220 pp.	\$4.00
July, 348 pp.	\$4.00
August, 350 pp.	\$4.00
September, 462 pp.	\$6.00
October, 490 pp.	\$6.00
November, 344 pp.	\$4.00
December, 390 pp.	\$5.00

SYMPOSIA REPRINTED from JGR

International Symposium on Electronic Distance-Measuring Techniques (144-page Symposium reprinted from the February 1960 issue)	\$3.50
Symposium on Sferics and Thunderstorm Electricity (102-page Symposium reprinted from the July 1960 issue)	\$3.50
Symposium on the Exosphere and Upper <i>F</i> Region (74-page Symposium reprinted from the September 1960 issue)	\$2.50
Scientific Effects of Artificially Introduced Radiations at High Altitudes (74-page Symposium reprinted from the August 1959 issue)	\$1.50
International Symposium on Fluid Mechanics in the Ionosphere (202-page Symposium reprinted from the December 1959 issue)	\$4.50

AMERICAN GEOPHYSICAL UNION

1515 Massachusetts Avenue, N.W., Washington 5, D. C.

To obtain back issues for 1958 and earlier years, write to
Walter J. Johnson, Inc., 111 Fifth Avenue, New York 3, New York.

Physicists

Hughes Research Laboratories offer an outstanding professional environment to Scientists and Engineers with an interest in basic and applied research. These exceptional facilities you can realize maximum professional growth through programs including:

ATMOSPHERIC PHYSICS

Thermionic energy conversion
Thermionic plasma devices

SURFACE PHYSICS

Low work function surfaces
Thermionic electron emission
Thermionic ion emission

SOLID STATE PHYSICS RESEARCH

Microwave and millimeter wave magnetic resonance
Optical and infrared spectroscopy
Solid State masers
Semiconductor physics
Radiation Physics

ION PROPULSION

The facilities of the Hughes Research Laboratories are located in Malibu, California, overlooking the Pacific Ocean—immediately adjacent to major academic institutions—with programs of academic support and participation. These facilities were specifically designed for effective research efforts with private offices and complete research laboratories. In this uniquely creative atmosphere, Hughes scientists are continually adding to their record of accomplishments in electronics and physics research.

Your inquiry may be directed in strict confidence to: Mr. D. A. Bowdoin, Hughes Research Laboratories, Malibu, California.

All qualified applicants will receive consideration for employment without regard to race, creed, color or national origin.

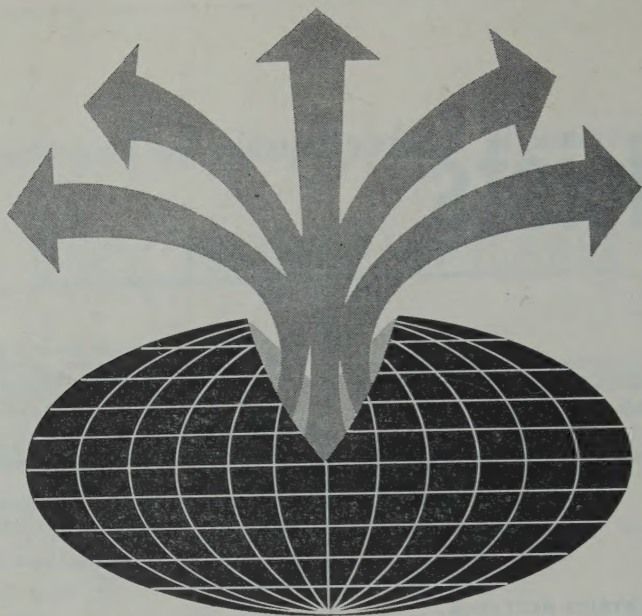
creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY
RESEARCH LABORATORIES



Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers



EARTH SCIENCES AT DRESSER ELECTRONICS, NATURALLY!

Since 1945 Dresser Electronics has been solving instrumentation problems associated with earth probes—first to find oil traps for petroleum geophysicists, now also to provide immediate answers to vital defense and security problems. A sample of present contracts—

for: U.S. GEOLOGICAL SURVEY (Major Crustal Studies Project)—Long-Range Recording Systems for VELA UNIFORM program.

DEPT. OF DEFENSE (Civil Defense)—an Automatic Radiation Monitoring System.

USAF (ARPA)—development and prototype of Unmanned Seismic Station.

US ARMY (Signal Supply Agency)—instruments and field studies relevant to seismic techniques for Missile Impact Location.

USAF (Systems Command/Hanscom Field)—development of VLF Electronic Seismometer with possible lunar probe application.

Dresser/SIE systems engineering is also at work in other projects throughout military, industrial and defense areas.

The unique approach of this systems-oriented company may solve your problem.

DRESSER ELECTRONICS

SIE

DIVISION

ONE OF THE DRESSER INDUSTRIES

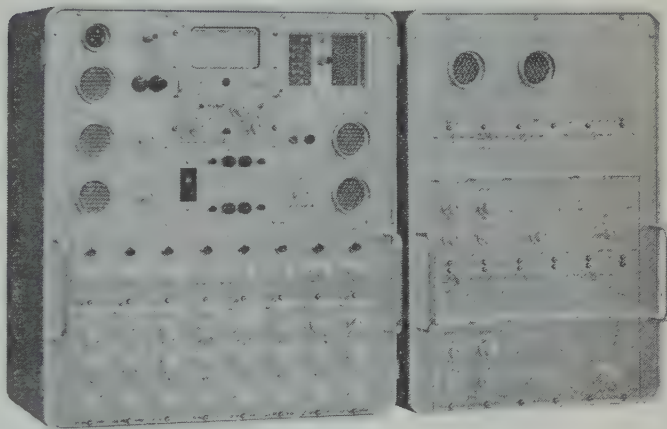
10201 Westheimer, Houston 42, Texas

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertiser

...YOU MAY BE ABLE RECORD SEISMIC DATA BEFORE POSSIBLE!

Performance of SIE's newest transistorized seismic instruments may let you perform seismic measurements up to now thought impossible. Here are two examples of the kind of equipment involved:

Transistorized Very-Low-Frequency Reflection Deep-Reflection Amplifier System. This system was originally developed and proven by use for the VELA UN-1 studies program being conducted by the Geological Survey. Transistorized, very-low-frequency amplifiers included in the system provide a new dimension for recording. Its low noise, wide dynamic frequency response to below 1 cps makes the TR-200 system ideal for all types of refraction and long range studies and records over longer distances with less noise!



PT-300 Ground Motion Amplifier For even more extended low-frequency response with outstanding signal-to-noise characteristics, SIE produced the PT300. Included are high-cut filters with provisions for either 18 or 36 db/octave slopes with cut-off at 9, 13, 18, 26, 37 or 100 cps. Frequency response extends from 100 cps down to 0.025 cps. An input sensitivity of 0.1 microvolt is provided while the total noise value is in the order of 10^{-8} volts!

For your problems call for a new product concerned with advanced seismic technology, the company who has continuously set the instrumentation standards of the industry—SIE

DIVISION

DRESSER ELECTRONICS

10201 Westheimer, Zone 42 / P. O. Box 22187, Zone 27 / Houston, Texas
PHONE: SUNset 2-2000 CABLE: SIECO HOUSTON TWX: HO-1185

Backed by 16 years of world leadership and world-wide service. Offices in Canada, Europe and Mexico—agents throughout the free world

GEOPHYSICAL MONOGRAPH SERIES

AMERICAN GEOPHYSICAL UNION

1515 MASSACHUSETTS AVENUE, N.W.

WASHINGTON 5, D. C., U.S.A.

Antarctica in the International Geophysical Year—Geophysical Monograph No. 1 (Publication No. 462, National Academy of Sciences—National Research Council); Library of Congress Catalogue Card No. 56-60071; 133 pp. and large folded map of the Antarctic, 1956, 7" x 10", \$6.00. Contains 16 pages by various American authorities on the Antarctic under the headings: General, Geographic and Meteorological, Geological and Structural, Upper Atmospheric Physics, and Flora and Fauna. Map (41" x 41") compiled by the American Geographical Society. Introduction by L. M. Gould.

Geophysics and the IGY—Geophysical Monograph No. 2 (Publication No. 590, National Academy of Sciences—National Research Council); Library of Congress Catalogue Card No. 58-60035; 210 pp., 1958, 7" x 10", \$8.00. Contains 30 papers by leading American authorities under the headings: Upper Atmospheric Physics, The Lower Atmosphere and the Earth, and The Polar Regions. Preface by Joseph Kaplan.

Atmospheric Chemistry of Chlorine and Sulfur Compounds—Geophysical Monograph No. 3 (Publication No. 652, National Academy of Sciences—National Research Council); Library of Congress Catalogue Card No. 59-60039; 129 pp., 1959, 7" x 10", \$5.50. Based on a symposium held jointly with the Robert A. Taft Sanitary Engineering Center of the U. S. Public Health Service in Cincinnati in November, 1957. Contains 23 papers (some as summaries) with discussion. Preface by James P. Lodge, Jr.

Contemporary Geodesy—Geophysical Monograph No. 4 (Publication No. 708, National Academy of Sciences—National Research Council); Library of Congress Catalogue Card No. 59-60065; 96 pp., 7" x 10", 1959, \$5.50. Based on a Conference held at Cambridge, Massachusetts, in December 1958 jointly by the AGU with the Smithsonian Astrophysical Observatory and the Harvard College Observatory. Contains 14 papers by leading authorities, with verbatim discussions on topics ranging from planetary-scale phenomena to microanalysis including hail formation and space navigation in the solar system. Edited by Charles A. Whitten and Kenneth H. Drummond.

Physics of Precipitation—Geophysical Monograph No. 5 (Publication No. 746, National Academy of Sciences—National Research Council); Library of Congress Catalogue Card No. 60-60010; 435 pp., 7" x 10", 1960, \$12.50. Based on a Conference held at Woods Hole, Massachusetts, in June 1959. Contains 48 papers by leading authorities, with verbatim discussions on topics ranging from planetary-scale phenomena to microanalysis including hail formation and precipitation control. Edited by Helmut Weickmann.

Postage is to be added to prices shown unless payment accompanies order. Quantity discounts (count each Monograph separately): 5-19 copies, 10%; 20-49 copies, 15%; 50 or more copies 20%.

Purchase Order

TO AMERICAN GEOPHYSICAL UNION

1515 Massachusetts Avenue, N.W., Washington 5, D. C., U.S.A.

Please enter our order for the following:

_____ copies of Geophysical Monograph No. 1, at \$6.00 *	\$ _____
_____ copies of Geophysical Monograph No. 2, at \$8.00 *	\$ _____
_____ copies of Geophysical Monograph No. 3, at \$5.50 *	\$ _____
_____ copies of Geophysical Monograph No. 4, at \$5.50 *	\$ _____
_____ copies of Geophysical Monograph No. 5, at \$12.50 *	\$ _____

☐ Payment of \$ _____ is enclosed.

☐ Please send invoice, adding postage charges.

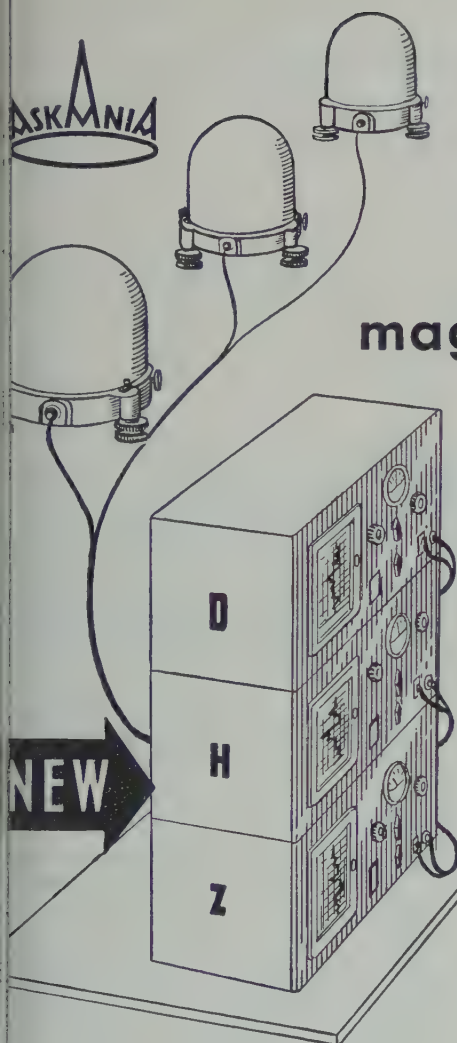
☐ Enter our standing order for _____ copies of subsequent Geophysical Monographs at the special prepublication rates, e.g., prepublication rate for Monograph No. 4 for non-members was \$4.00, payment in advance, or \$4.75 (plus postage) on invoice.

* List price is net for quantities up to four; see above for discounts on quantity purchases. Special discounts to members.

Typed name _____ Signature _____

Address _____

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers



REMOTE RECORDING of magn. variations by three MAGNETO- GRAPHS TYPE Gm1

- fibre suspension for all 3 magnet systems
- photoelectric sensing elements
- thermostats
- scale value adjustable between 3 and 20 gammas per mm by electric means

for detailed information

ASKANIA-WERKE

Division of Continental Elektroindustrie AG

Branch

4913 Cordell Avenue

Bethesda 14, Maryland

43RD ANNUAL MEETING
 AMERICAN GEOPHYSICAL UNION
 APRIL 25 through 28, 1962
 WASHINGTON, D. C.

*your participation
 is welcomed*

If you want to present a paper
 send the abstract BEFORE DEADLINE DATE to

<i>Hydrology</i>	<i>before January 2, 1962, to Harry E. Schwarz, 315 Ladson Road, Silver Spring, Md.</i>
<i>Meteorology</i>	<i>before January 9, 1962, to William Hiatt, US Weather Bureau, Washington 25, D. C. (abstract plus comprehensive summary)</i>
<i>Geodesy</i>	<i>before February 1, 1962, to Erwin Schmid, US Coast & Geodetic Survey, Washington 25, D. C.</i>
<i>Seismology</i>	<i>before February 1, 1962, to Benjamin F. Howell, Jr., Geophysical Laboratory, Mineral Sciences Bldg., Pennsylvania State Univ., University Park, Penna.</i>
<i>Oceanography</i>	<i>before February 1, 1962, to Arthur E. Maxwell, 5627 Potomac Ave., N.W., Washington 16, D. C.</i>
<i>Volcanology, Geochemistry, and Petrology</i>	<i>before February 1, 1962, to David R. Wones, US Geological Survey, Washington 25, D. C.</i>
<i>Geomagnetism and Aeronomy</i>	<i>before February 1, 1962, to J. Hugh Nelson, US Coast and Geodetic Survey, Washington 25, D. C.</i>
<i>Tectonophysics</i>	<i>before February 1, 1962, to Mrs. Iris Borg, Box 579, Livermore, California.</i>
<i>Planetary Sciences</i>	<i>before February 1, 1962, to Robert Jastrow, Goddard Institute for Space Studies, 475 Riverside Dr., New York 28, N. Y.</i>
<i>General Program</i>	<i>to Leroy R. Alldredge, 10500 Royal Rd., Silver Spring, Md.</i>

Be sure to

send 3 copies, including ribbon copy, of the abstract
 double or triple space
 make it concise—about 200 words

use correct format: John Doe (Physics Dept., Iowa State University, Ames, Iowa) Short-Period Micropulsations—The geographical extent,

see the December *Transactions* for complete details on the meeting

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

Journal of GEOPHYSICAL RESEARCH

IE 66

NOVEMBER 1961

No. 11

Formation of the Geomagnetic Storm Main-Phase Ring Current¹

A. J. DESSLER, W. B. HANSON, AND E. N. PARKER²

*Lockheed Missiles and Space Company
Palo Alto, California*

Abstract. It is suggested that hydromagnetic waves generated by the impact of solar plasma on the geomagnetic field may form shock waves in the magnetosphere, thus providing a mechanism for establishing the diamagnetic main-phase ring current. These shock waves should develop on the night side of the earth and heat the ambient protons (which constitute the thermal protonosphere) to approximately the hydromagnetic-wave velocity (of the order of 10 km/sec). The transfer of hydromagnetic-wave energy to the protons stresses the geomagnetic field and produces the geomagnetic-storm main phase; i.e., the kilovolt-energy protons form a diamagnetic current. The bombardment of the upper atmosphere by energetic hydrogen ions from the decaying ring current and the possible change in the decay time-constant of the ring current through the sunspot cycle are discussed.

INTRODUCTION

The main phase of a geomagnetic storm is the period of the storm during which the geomagnetic field strength is below the normal un disturbed value at low and temperate latitudes. This hypothesis, first put forth by *Singer* [1957], that the main-phase decrease is due to energetic protons trapped in the geomagnetic field is generally accepted. It was initially assumed these protons were from solar plasma injected into the geomagnetic field [*Singer*, 1957; *Forster and Parker*, 1959]. Two independent experiments indicated that the main phase is caused by kilovolt-energy protons in a ring current belt somewhere in the vicinity of 4 earth radii ($4 R_E$) from the center of the earth [*Desler and Parker*, 1959]. The approximate position of the ring current was inferred from (1) the observed decrease of the vertical intensity at magnetic latitudes of about 60° and (2) the observed 1-day main-phase recovery time, i.e., the

time for the ring current to decay. This location for the main-phase ring current is consistent with the magnetic-storm observations of Explorer VI [*Smith and Rosen*, 1960].

Zodiacal-light measurements [*Blackwell and Ingham*, 1961] and our interpretation of the Explorer VI radiation measurements [*Arnoldy, Hoffman, and Winckler*, 1960; *Fan, Meyer, and Simpson*, 1960; *Rosen, Farley, and Sonett*, 1960] indicate that the solar wind is much weaker than has been previously thought, so that it does not penetrate the geomagnetic field to as close as $4 R_E$ geocentric distance. For example, if the solar wind had pushed in as close as $4 R_E$ during the magnetic storm of August 15-18, 1959, when our interpretation of Explorer VI data indicated a diamagnetic ring current at $4 R_E$, the Van Allen radiation beyond approximately this distance could not have remained trapped (see *Dessler and Karplus* [1961] for a fuller discussion). However, the radiation detectors indicated the presence of trapped radiation all the way out to apogee (about $7 R_E$). Thus, we doubt now that the particles at $4 R_E$, which produce the geomagnetic-storm main phase, are of direct solar

¹Presented at the meeting of The American Geophysical Society, June 22-24, 1961, Mexico City.
²Permanent address: Enrico Fermi Institute for Nuclear Studies, University of Chicago, Chicago, Illinois.

origin, and we must look elsewhere for the origin of the trapped particles. An attractive alternative to direct injection is to assume that charged particles already present in the geomagnetic field (i.e., the particles which constitute the protonosphere) are accelerated in the vicinity of $4 R_E$ and beyond.

The depression of the horizontal component of the geomagnetic field at the surface of the earth is proportional to the total energy of the trapped particles, and about 10^{15} joules of trapped particle energy is needed to give the typical 100- γ reduction of a big storm. The proton density at $4 R_E$ is of the order of 10^2 protons/cm³ [Smith, 1960; Johnson, 1961], so that the total energy requirement suggests an individual proton energy of the order of 1 keV (500 km/sec). To account for the main phase of the geomagnetic storm, the protons must be accelerated or heated to 1-keV energy during the active phase of the storm. The present paper suggests that the heating is hydromagnetic in nature, in a manner similar to the heating of the solar corona.

It is widely believed that the solar corona is heated by the dissipation of hydromagnetic or sound shock waves. The particle collision frequency is negligibly small in the corona com-

pared with the shock-wave rise time. No quantitative theory has been put forth to describe energy dissipation in a collisionless shock. However, observations clearly show that the corona is being continuously heated (see a review of the development of coronal-heating theories in the appendix of Parker [1960]); we will assume that the earth's protonosphere is heated by a shock dissipation mechanism similar to that operating in the solar corona.

THE MODEL

To illustrate how hydromagnetic shocks may generate a proton ring current, we note that the outer geomagnetic field is distorted by the solar wind, perhaps in a tear-drop shape, as suggested by Johnson [1960]. The essential feature is that the geomagnetic field is unsymmetric, extending farther from the earth on the night side than on the day side. Experimental evidence indicates that, at a time of unusually low magnetic activity, the boundary between the solar wind and the geomagnetic field near local noon occurs at about $14 R_E$ [Sonett, Judge, Sims, and Kelso, 1960].

It has been pointed out that large-amplitude hydromagnetic waves can be generated by the impact of solar plasma on the geomagnetic field [Dessler, 1958; Dessler and Parker, 1959]. These hydromagnetic waves are generated where the solar wind comes in contact with the geomagnetic field, so that their source extends around the sunward side. The breadth of the source indicates that, even on the night side, there will be no significant diminution of wave amplitude resulting from geometrical attenuation. The geomagnetic field strength is approximately doubled at the interface on the sunward side by solar-wind pressure. On the night side the field is more nearly dipolar, so that the geocentric distance at which $\Delta B/B = 1$ will be less on the night side than on the day side for a given wave amplitude ΔB . The distortion of the surfaces of constant B is illustrated in Figure 1. For example, if $\Delta B/B = 1$ at $6 R_E$ on the sunward side, then $\Delta B/B$ will again equal unity at approximately $6/(2)^{1/2} = 4.8 R_E$ on the night side. Beyond this distance, $\Delta B/B < 1$ since B falls off roughly as $1/r^3$; closer to earth, $\Delta B/B < 1$ since B rapidly increases. This model is then analogous to the solar corona where hydromagnetic waves move outward

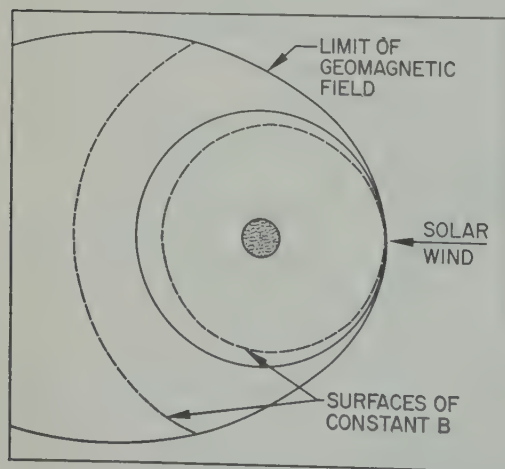


Fig. 1. A sketch showing the solar-wind-geomagnetic field interface in the equatorial plane when the solar wind has pushed to within 6 earth radii. A solid-line circle at $6 R_E$ is shown, together with dashed lines representing surfaces of constant B . Note that the value of B at the subsolar point on the interface is the same as the value of B at the antisolar point at a distance of only $4.8 R_E$.

into a region of monotonically decreasing strength until $\Delta B/B \geq 1$. We will see in the next section that hydromagnetic pressure increases rapidly with $\Delta B/B$, so that toward the night side that we expect the hydromagnetic heating in the geomagnetic field.

HYDROMAGNETIC HEATING

The theory of the hydromagnetic heating of a ionized gas has been developed because of its importance in maintaining the solar corona. Heating of a tenuous gas through dissipation of hydromagnetic waves is based on the theory of both longitudinal and transverse hydromagnetic waves to steepen their fronts. The rate of steepening of hydromagnetic waves depends upon their relative amplitude and the compressibility of the medium. The rate of steepening is significant when the wave amplitude approaches B , and when the medium is incompressible, i.e., when the gas pressure $p \ll B^2/2\mu_0$. The steepening increases until it is limited by dissipation. The dissipation increases as the square of the gradient when there are collisions (i.e., when viscosity and resistivity are properties of the medium). In the absence of collisional dissipation, dynamical instabilities take over the dissipative process. The heating is self-regulating, raising the temperature of sound to a value comparable to the wave amplitude of the wave, at which level steepening is limited by declining compressibility.

A compressional wave, whether it propagates parallel or perpendicular to a magnetic field, will steepen its front because, among other things, the speed of propagation is higher in the compression than in the rarefaction. The theory of steepening of a longitudinal wave (an acoustic wave) in a field-free gas, or in the presence of a uniform magnetic field in the direction of propagation, may be found in the literature (Lighthill, 1956; Taylor and Maccoll, 1957) for a polytrope gas, for which $p \propto \rho^\alpha$, where α is the polytrope index for the gas. The steepening time for a compressional wave propagating across a magnetic field has been calculated by Petschek [1958].

A transverse hydromagnetic wave propagating in a uniform magnetic field through an incompressible medium is nondispersive. A plane

wave that is exactly circularly polarized is nondispersive, even when the medium is highly compressible [Ferraro, 1955], but for any other case the wavefront tends to steepen when there is compressibility. The steepening time for transverse hydromagnetic waves has been worked out by Montgomery [1959] for the case where the compressibility of the gas is large, $p \ll B^2/2\mu_0$. The steepening time for a longitudinal wave is proportional to $B/\Delta B$, whereas for the transverse wave it is proportional to $(B/\Delta B)^2$. The difference arises from the fact that in the former case the pressure change is first order in ΔB , whereas in the latter it is second order in ΔB .

It follows immediately from the above work that the sufficient conditions for rapid steepening of a hydromagnetic wave propagating in a uniform magnetic field B are a compressible medium and $\Delta B \sim B$. When these conditions are satisfied, the steepening time t_s may be shown to be of the order of the wave period T . For the case of a wave propagating into the tail of the geomagnetic field, t_s will be less than T because:

1. ΔB must increase with radial distance if there is no dissipation. This statement follows from conservation of energy flux $(\Delta B^2/\mu_0)V_{hm}$ where V_{hm} is the hydromagnetic wave velocity; since V_{hm} decreases radially outward approximately as $1/r^2$ [Dessler, Francis, and Parker, 1960], ΔB must increase directly proportional to r .

2. B decreases as $1/r^2$. Therefore $\Delta B/B$ increases about as r^4 . This increase in $\Delta B/B$ with radial distance is so rapid that $\Delta B/B$ will increase significantly in less than a wavelength for wave periods greater than about 30 seconds; for example, between 4 and 5 R_E , $\Delta B/B$ will increase by $(5/4)^4 \approx 2.5$. The wavelength for a 30-second-period wave at 4.5 R_E is about 10^4 km or $1\frac{1}{2}$ R_E . In this example, $\Delta B/B$ is more than doubled in approximately $2/3$ of a wavelength.

The self-regulating aspect of hydromagnetic shock heating may be seen from the following discussion. Starting with a cold gas ($p \ll B^2/2\mu_0$), the interaction between the solar wind and the geomagnetic field produces both longitudinal and transverse hydromagnetic waves. The propagation velocity of both types of waves within most of the magnetosphere is $B/(\mu_0\rho)^{1/2}$; we would expect the energies in the two modes to be of

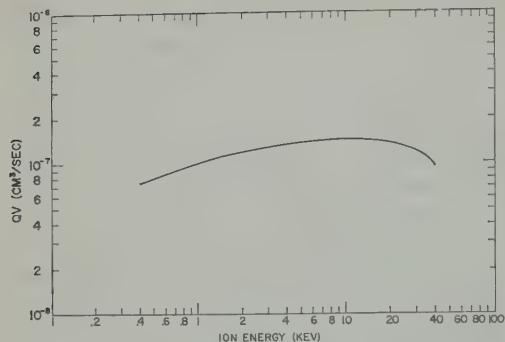


Fig. 2. The product of the charge-exchange cross section Q , times the proton velocity V , plotted against proton energy (based on the data of Fite, Stebbings, Hummer, and Brackmann [1960]). The quantity $(nQV)^{-1}$, where n is the hydrogen atom concentration, represents the effective proton-ring-current lifetime. Note that, for a given value of n , the proton lifetime is nearly constant over a wide range of energy.

the same order. The characteristic steepening and dissipation time is less than the individual wave period T , so that heating proceeds rapidly. We postulate here that the energy should go principally into the ions. The vigorous supply of hydromagnetic waves present during the active phase of the storm is sufficient to raise the ion temperature rather quickly. The heating goes on until the ion pressure becomes comparable to the field pressure and thus reduces the compressibility of the medium. This relative decrease of compressibility will increase the characteristic dissipation time. Thus, heating will diminish when p becomes comparable to $B^2/(2\mu_0)$. In addition, a cooling effect will occur if p should momentarily exceed $B^2/(2\mu_0)$. The field could then no longer contain the ions, which would expand (or escape through various dynamical instability processes) and be lost from the region. Altogether, then, we expect the heating to proceed to the point where p is of the same order as $B^2/(2\mu_0)$, and no further. Since $\Delta B \sim B$, the speed of sound is comparable, in order of magnitude, to the velocity amplitude of the individual waves—the *Mach 1 effect*. With $\Delta B/B \sim 1$, we expect that the thermal velocity of the ambient ions will be raised to the hydromagnetic velocity, which is of the order of the required 500 km/sec (corresponding to about 1-kev energy) in the general vicinity of $4 R_E$ [Dessler, Francis, and Parker, 1960]. The hydro-

magnetic wave velocity may increase by about a factor of 3 during a magnetic storm because of an observed decrease in the plasma density of the protonosphere [Carpenter, 1961]. Such an occurrence would not change any of the arguments or affect any of the conclusions reached in this paper. The heating is not effective much closer than about $4 R_E$ because B increases inward so rapidly that $\Delta B/B$ is soon much less than unity. Thus, we expect that ambient protons beyond about $4 R_E$ will be heated by hydromagnetic shock waves during the active phase of a geomagnetic storm until the proton energy density approaches the local magnetic-field energy density.

A solar-wind density of 50 protons/cm³ moving with a velocity of 10^3 km/sec delivers about 4×10^{-2} watt/m² to the geomagnetic field, so that over a cross section of radius $4 R_E$ the energy incident is 8×10^{13} joules/sec. When we assume that a significant fraction of this energy goes into hydromagnetic-wave generation, we find that there is ample wave energy during the active phase of the storm to generate the 10^{15} joules necessary to produce the main phase.

The hydromagnetic-heating mechanism proposed in this paper transfers an important amount of energy only to the protons; for conditions applicable to the protonosphere, the electron energy remains substantially unaffected by the passage of a hydromagnetic shock wave as the hydromagnetic velocity corresponds to only $\frac{1}{2}$ ev for an electron. An independent argument is here advanced to show that energetic protons rather than electrons constitute the main-phase ring current. We note that there is satisfactory agreement between the time constant for decay of the proton ring current, the charge exchange with the telluric hydrogen corona and the observed magnetic-storm recovery time [Dessler and Parker, 1959]. The charge-exchange time constant is rather independent of the proton energy, as shown in Figure 2. If the energetic particles forming the ring current were electrons rather than protons, the recovery time would be too long, as shown by the following estimate of the interaction time between electrons and the outer atmosphere. Electrons are subject to interactions with the groups of particles which could cause the eventual decay of the ring current. The interaction

the fast-trapped electrons themselves cause a randomizing of their pitch angles, would feed particles into small pitch and cause their penetration to low altitudes where they would be lost. The elastic collisions between the electrons and protons would affect the ring-current energy directly, but collisions are about as effective as the electron-electron collisions in randomizing the pitch angles. Also, the inelastic collisions with the neutral hydrogen present would cause a direct loss of energy from the ring current.

An estimate of the time constant for the first mechanism can be obtained from the formula for the deflection time t_D required to produce a 90° change in the direction of a charged particle moving in a gas of similar particles (Carpenter, 1956),

$$t_D = \frac{11.4 A^{1/2} T^{3/2}}{n \ln \Lambda} \quad (1)$$

A is the atomic weight of the particles (36 for electrons), T is the temperature in degrees Kelvin, n is the particle concentration, and Λ is a slowly varying function of n and T which is approximately 30 for the values in this paper. Thus, for electron-electron scatter-

$$t_D = 1.1 \times 10^4 \frac{E^{3/2}}{n} \text{ sec} \quad (2)$$

E is in electron volts and n is the number of electrons/cm³. For a particle energy of 1 kev and a concentration of 200 electrons/cm³, we have $t_D = 1.8 \times 10^6$ sec, i.e., approximately 3 days. Assuming that protons have the same scattering effectiveness as electrons, this time would be reduced to about 10 days. If the electron concentration were reduced during magnetic storms [Carpenter, 1961] then the individual particle energy would be increased. Both these effects would tend to increase the electron loss time given by equation 2.

Another process which must be considered is inelastic collisions of 1-kev electrons with hydrogen atoms. The average energy lost in a randomizing collision is approximately 50 ev, about 10 of these collisions per electron are needed to reduce the ring-current energy by a factor of 2. If we take 2×10^{-17} cm² as the

average ionization cross section Q_i [Fite and Brackmann, 1958], then the time required for 10 collisions is approximately

$$t_i = \frac{10}{n(H) Q_i \bar{v}_e} = 3 \times 10^8 / n(H) \quad (3)$$

where $n(H)$ is the atomic-hydrogen concentration. From Figure 2 we see that the decay time for 1-kev protons by charge exchange is $10^7 / n(H)$. The energy-loss time for the electrons is about 30 times larger than the charge-exchange time for protons, with the same atomic-hydrogen concentration. Thus, our calculations indicate that at least 1 week is required to remove a trapped 1-kev-electron ring current; the time constant for electrons is too long to match the observed ring-current decay (of the order of 1 day).

PHYSICAL CONSEQUENCES OF A PROTON RING CURRENT

High-energy atomic-hydrogen flux from decaying proton ring current. On the basis of the shock-wave heating mechanism proposed in this paper, only the protons would be significantly affected by the shock-wave dissipation. The decay of the proton ring current by charge exchange with hydrogen atoms gives rise to a flux of fast neutral hydrogen atoms (~ 1 kev) penetrating below the base of the exosphere. The fraction f of the fast hydrogen atoms which arrive at the top of atmosphere for a ring current at $4 R_E$ is given approximately by

$$f = \frac{\pi R_E^2}{4\pi (4R_E)^2} = 1/64 \quad (4)$$

Since 10^{31} protons are lost in a time of the order of 10^5 seconds, this gives rise to a flux ϕ of fast hydrogen atoms at the base of the exosphere of

$$\begin{aligned} \phi &= \frac{10^{31}}{10^5 \times 64} \times \frac{1}{4\pi R_E^2} \\ &= 3 \times 10^5 \text{ atoms/cm}^2\text{-sec} \end{aligned} \quad (5)$$

Observations of this flux during the main phase would be strong evidence in support of a proton ring current. These observations should be made on the night side, however, since there could be a large background hydrogen-atom flux on the

base of the exosphere on the day side caused by charge exchange between the solar wind and the escaping-hydrogen geocorona.

Change in ring-current decay time with sunspot cycle. Johnson [1961] has pointed out that the radiation source responsible for atomic-hydrogen formation (~ 2000 Å) does not change appreciably through the solar cycle, whereas the radiation (< 1000 Å) which determines the temperature of the exosphere is probably less intense during sunspot minimum than during sunspot maximum; thus, in order to match the escape and formation rates, it follows that the geocorona must be more dense near sunspot minimum than near sunspot maximum. At $4 R_E$, this change in the hydrogen concentration amounts to a factor of about 3 [Johnson, 1961, p. 17] over the sunspot cycle. If the ring current decays by charge exchange with the hydrogen geocorona, the time constant of this decay should be about three times shorter during sunspot minimum than during sunspot maximum (assuming that the position of the ring current does not vary appreciably over the solar cycle). An examination of existing magnetic-storm magnetograms could thus provide a direct test of this hypothesis for the removal of the diamagnetic ring current.

CONCLUSIONS

1. It is possible for the geomagnetic-storm main phase ring current to be formed by heating the normal protonosphere by hydromagnetic shock waves.

2. It is unlikely that the ring current is composed of energetic electrons.

3. If our conclusion that the ring current is formed of kilovolt-energy protons that are removed by charge exchange is correct, the decay time of the ring current should be less during years of sunspot minimum than during sunspot maximum.

Note added in proof. S. Matsushita (private communication) has found a variation in magnetic storm recovery time with solar activity. The recovery time appears to be about three times shorter during sunspot minimum than during sunspot maximum.

Acknowledgments. We wish to thank Miss Janet Gordon for her assistance in the preparation of this manuscript.

This work was supported in part by the Geophysics Research Directorate of the Air Force Cambridge Research Laboratories under contract AF 19(604)-7989, and in part by the Goddard Space Flight Center of the National Aeronautics and Space Administration under contract NAS 5-657.

REFERENCES

- Arnoldy, R. L., R. A. Hoffman, and J. R. Winckler, Observations of the Van Allen radiation regions during August and September, 1959, Part I, *J. Geophys. Research*, **65**, 1361-1376, 1960.
- Blackwell, D. E., and M. F. Ingham, Observation of the zodiacal light from a very high altitude station, *Monthly Notices Roy. Astron. Soc.*, **12**, 113-155, 1961.
- Carpenter, D. L., Whistler data on the change of base level ionization in the whistler medium following magnetic storms, URSI Meeting, Washington, D. C., May 1-4, 1961.
- Dessler, A. J., Large-amplitude hydromagnetic waves above the ionosphere, *Phys. Rev. Letters*, **1**, 68-69, 1958; and *J. Geophys. Research*, **6**, 507-511, 1958.
- Dessler, A. J., and E. N. Parker, Hydromagnetic theory of geomagnetic storms, *J. Geophys. Research*, **64**, 2239-2252, 1959.
- Dessler, A. J., and R. Karplus, Some effects of diamagnetic ring currents on Van Allen radiation, *J. Geophys. Research*, **66**, 2289-2295, 1961.
- Dessler, A. J., W. E. Francis, and E. N. Parker, Geomagnetic storm sudden-commencement risetimes, *J. Geophys. Research*, **65**, 2715-2719, 1960.
- Fan, C. Y., P. Meyer, and J. A. Simpson, Trapped and cosmic radiation measurements from Explorer VI, *Space Research*, edited by H. Kallmann Bijl, North Holland Publishing Company, Amsterdam, 1195 pp., 1960.
- Ferraro, V. C. A., Hydromagnetic waves in a radiationized gas and galactic magnetic fields, *Proc. Roy. Soc., London*, **A233**, 310-318, 1955.
- Fite, W. L., and R. T. Brackmann, Collisions of electrons with hydrogen atoms, I, *Ionization Phys. Rev.*, **112**, 1141-1151, 1958.
- Fite, W. L., R. F. Stebbings, D. G. Hummer, and R. T. Brackmann, Ionization and charge transfer in proton-hydrogen atom collisions, *Phys. Rev.*, **119**, 663-668, 1960.
- Johnson, F. S., The gross character of the geomagnetic field in the solar wind, *J. Geophys. Research*, **65**, 3049-3051, 1960.
- Johnson, F. S., Structure of the upper atmosphere, chap. 1, *Satellite Environment Handbook*, edited by F. S. Johnson, Stanford University Press, Stanford, California, 155 pp., 1961.
- Lighthill, M. J., Viscosity effects in sound waves of finite amplitude, *Surveys in Mechanics*, edited by G. K. Batchelor and R. M. Davies, Cambridge University Press, 250-351, 1956.
- Montgomery, D., Development of hydromagnetic shocks from large amplitude Alfvén waves, *Phys. Rev. Letters*, **2**, 36-37, 1959.

- E. N., The hydromagnetic theory of solar secular radiation and stellar winds, *Astrophys. J.*, **132**, 821-866, 1960.
- K. H. E., Aerodynamic dissipation, *Revs. Modern Phys.*, **30**, 966-972, 1958.
- A., T. A. Farley, and C. P. Sonett, Soft ion measurements on Explorer VI earth satellite, *Space Research*, edited by H. K. Kallipour, North Holland Publishing Company, Amsterdam, 1195 pp., 1960.
- S. F., A new model of magnetic storms and substorms, *Trans. Am. Geophys. Union*, **38**, 175-185, 1957.
- E. J., and A. Rosen, The scientific results of the satellite Explorer VI, *Ballistic Missiles and Space Technology*, Vol. III, Academic Press, New York, 450 pp., 1960.
- Smith, R. L., The use of nose whistlers in the study of the outer ionosphere, *Tech. Rept. no. 6*, OSR contract AF 18(603)-126, Stanford Electronics Laboratories, 1960.
- Sonett, C. P., D. L. Judge, A. R. Sims, and J. M. Kelso, A radial rocket survey of the distant geomagnetic field, *J. Geophys. Research*, **65**, 55-68, 1960.
- Spitzer, L., Jr., *Physics of Fully Ionized Gases*, Interscience Publishers, New York, 105 pp., 1956.
- Taylor, G. I., and J. W. Maccoll, *Aerodynamic Theory*, edited by W. F. Durand, California Institute of Technology, Pasadena, **3**, 210-217, 1943.

(Manuscript received August 7, 1961; revised August 30, 1961.)

Spectrum of Hydromagnetic Waves in the Exosphere

GORDON J. F. MACDONALD

*Goddard Space Flight Center¹
National Aeronautics and Space Administration
Washington 25, D. C.*

Abstract. A disturbance in the exosphere generates waves in three partially separable modes. The modes are described by considering the vorticity about a line of force, the two-dimensional divergences of velocity in the plane perpendicular to the line of force, and the component of velocity along the line of force. The propagation of vorticity is one-dimensional and there is no geometrical attenuation. It is suggested that this mode is associated with the sudden commencement of a magnetic storm. The propagation of the longitudinal component velocity is almost one-dimensional, but it is heavily damped at high frequencies. In a gravitational field, the medium is no longer uniform, and at low frequencies the modes are coupled in a complicated way. The influence of gravity is characterized by the frequency of oscillation of a small parcel of fluid when displaced from its zero-order position in the gravitational field. For parallel magnetic and gravitational fields the vorticity mode is still separable, and gravity leads to anisotropic dispersion in the other modes. The propagation at low frequencies is no longer cylindrically symmetrical if the lines of force are at an arbitrary angle to the direction of gravity. Detailed investigations of the influence of the gravitational field on hydromagnetic wave propagation is carried out for the case where the sound velocity is small compared to the Alfvén wave velocity.

The principal aim of the study is an attempt to provide a background for the separation of source effects from those of the medium as an aid to the spectral study of magnetic records. The relative proportion of the three modes generated by a specific source is estimated. Consideration of the regions of allowed solutions in frequency-wave number space show that the vorticity mode will have a narrow spectrum. It is suggested that high-frequency micropulsations are associated with vorticity mode propagation. Examples of vorticity generating sources include the turbulent region in the vicinity of a finite amplitude wave, the outgassing of a rocket motor, and a wake of a satellite in the lower exosphere. Energy may be trapped within the exosphere in various free vorticity mode oscillations. The free periods of these oscillations are comparable to the periods observed in the low frequency micropulsations.

The limiting conditions for the propagation of the various modes in the exosphere are considered. Because of the long-range nature of the Coulomb forces, the charged particle may travel distances short compared with its mean free path and still exchange a substantial proportion of its momentum. For representative models of the exosphere, disturbances at several earth radii must have a length scale large compared with several kilometers for the propagation of the longitudinal component of velocity. The limiting conditions on the vorticity and transverse modes involve the ratio of the length scale of the disturbance to the Larmor radius. These modes are propagated if the disturbances have length scales of some hundreds of meters in the outer regions of the exosphere.

INTRODUCTION

The theory of magnetic storms has undergone remarkable development in the past few years. Early as 1896 Birkeland proposed that magnetic storms were the result of streams of particles emitted from the sun and guided towards earth by the earth's magnetic field. Chapman and Ferraro [1931; 1933; 1940] elaborated the particle theory. The theory presumes that

the space between the earth and the sun is a vacuum and that the induced currents in the plasma are propagated with the speed of light. Investigations of whistlers [Storey, 1953] and of zodiacal light [Behr and Siedentopf, 1953] indicated that the interplanetary space contains a resident ionized gas with a density of a few hundred electrons per cubic centimeter. A recent investigation of the zodiacal light by Blackwell and Ingham [1961] limits the density to less than 120 electrons/cc in the outer corona, whereas further whistler studies suggest a comparable value for the upper regions of the

¹Present address: Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California.

exosphere. The presence of a resident interplanetary ionized gas requires that the propagation of a solar stream into the earth's field be treated as a problem in hydromagnetics [Dessler, 1958]. Dessler and Parker [1959] have undertaken a detailed investigation of a hydromagnetic theory of geomagnetic storms. (An alternative development is due to Piddington [1959; 1960a, b].) The magnetic storm results from fluctuations in the intensity of particles streaming from the sun. Disturbances are propagated through the exosphere as hydromagnetic waves. On reaching the base of the ionosphere the hydromagnetic signals are transmitted through the atmosphere as ordinary electromagnetic waves and at the speed of light. In the lower ionosphere, with its high concentration of neutral particles, the propagation is complicated and may take the form of a diffusing electromagnetic wave [Fejer, 1960; Francis and Karplus, 1960]. The hydromagnetic theory of geomagnetic storms thus proposes that the magnetic signal is transferred through the exosphere at the local Alfvén wave velocity and through the un-ionized atmosphere at the speed of light.

Dessler and Parker, in their development of the hydromagnetic theory, emphasize the characteristic times associated with various events in a magnetic storm. Dessler, Francis, and Parker [1960] show that the time taken for hydromagnetic waves to pass from the solar plasma front to ground level is on the order of a few minutes. This prediction is in accord with the observed rise time of a sudden commencement. (The simultaneity of the world-wide increase in intensity of the horizontal component is open to question because varying standards of timing are used by geomagnetic observatories. Williams [1960] finds that sudden commencements are registered first in the high and middle latitudes.)

The present study is an attempt to provide a partial background for the separation of source effects from those of the medium as a prelude to the spectral analysis of magnetic records. A major limitation is the neglect of all dissipation.

Although much progress has been made in determining the time characteristics of geomagnetic storms, not as much effort has been expended in a study of the frequency characteristics of geomagnetic variations. A number of peridiogram, or zero-crossing, analyses have been made on geomagnetic records. Because of the noise level

and the complicated spectra, neither method lead to reliable spectral estimates. Perhaps the principal reasons for the lack of spectral estimates are the analog recording, usually used with magnetometers and earth current coils, and the difficulty in going from an analog to a digital record.

The sharp nature of the onset of a magnetic storm indicates a spectrum relatively rich at high frequencies (cycles per second to cycles per minute). The spectrum associated with the sharp onset may contain evidence of the source. Dessler and Parker [1959] suggest that there is a high frequency cutoff fixed by the Larmor frequency at the source. The long time-scale associated with the main phase of the magnetic storm suggests that the spectrum will also contain energy at much lower frequencies (cycles per hour). Furthermore, instruments of sufficient sensitivity have established the existence of micropulsations in the geomagnetic field. These are commonly associated with the onset of magnetic storms and show sinusoidal time variations with periods ranging from a few tenths of a second to a hundred seconds [Benioff, 1960; Jacobs and Sin, 1960; Troitskaya, 1961]. It is tempting to describe the narrow band of frequencies associated with the low-frequency micropulsations to the fact that they represent a resonant mode of the exosphere. This interpretation is strengthened by their presence during periods of relative magnetic quiet.

In the presence of a magnetic field a disturbance is propagated through a uniform, electrically conducting fluid by three wave fronts traveling at three different speeds [Herlofson, 1950; Friedrichs and Kranzer, 1958; Bazer and Fleischman, 1959; and Grad, 1959]. The three modes are illustrated in the case where the Alfvén wave velocity is large compared with the sound velocity. The vorticity about the line of force is propagated one-dimensionally at the Alfvén wave velocity. The component of velocity along the line of force is propagated at the sound velocity and the propagation is almost one-dimensional. The two-dimensional divergence in the plane perpendicular to the line of force, together with the longitudinal component of the magnetic field, is propagated almost spherically at the Alfvén wave velocity. For convenience we will label the one-dimensional vorticity mode as the V mode. The nearly c

onal sound wave is termed the pressure, P , mode, while the spherically spreading mode is termed the transverse, or T , mode. If the wave velocity is much less than sound velocity, the V mode remains one-dimensional, the P mode is spherically spreading, and the T mode is nearly one-dimensional. It should be emphasized that, in general, the solutions of the equations cannot be resolved into three modes characterized by a velocity, and the wave fronts interact to give a complicated wave form.

Part 1 presents a review of the theory of propagation of waves in an anisotropic inhomogeneous medium. Dissipation is ignored. The dynamic equations lead to dispersion relations which for a fixed frequency determine slowness surfaces) in wave number space. A slowness surface, in turn, fixes the geometry of a propagating wave. The concept of group velocity is particularly important, since the direction of propagation of the phase of the wave and the direction of the energy propagation need not be the same. The J.W.K.B. approximation in the propagation of waves in an inhomogeneous medium is considered, and the limitations of the method are stressed. The method yields results which are valid only if the wave length of the disturbance is short compared with the scale of the variable properties. For many problems in hydromagnetic wave propagation in the exosphere this holds true, but the approximation is not as good for interesting low-frequency waves. Despite the limitations of the approximation, a qualitative picture of low-frequency waves may be obtained.

Parts 2 and 3 treat the propagation of the low-frequency hydromagnetic waves with the aid of the diagnostic diagrams introduced by *Eckart* [1958] in his study of classical gravity waves. A diagnostic diagram is a projection in the wave-number-frequency space of the regions of allowed modes of propagation. At low frequencies it is possible for gravity to modify the various hydromagnetic modes. In Part 2 the case of the combined magnetic and gravitational fields is considered. The slowness surfaces and surfaces of constant phase are determined for various frequencies. Gravity leads to anisotropic dispersion. The influence of gravity is negligible at low frequencies (period much less than several seconds).

There is a balance of particle and magnetic

pressure in the P mode propagation. A decrease in particle pressure is accompanied by an increase in the local magnetic field. An increase in the magnetic field is accompanied by a transfer of energy from the acoustic field, and the particle pressure decreases. The ratio of the variation in particle pressure to the variation in magnetic pressure is determined by the square of the ratio of sound to Alfvén wave velocity in the T mode. The variation of particle pressure during the passage of a T wave in the exosphere is small. The V wave does not alter the thermodynamic state of the medium.

Part 3 considers the case of perpendicular gravitational and magnetic fields. The propagation is not cylindrically symmetrical if the lines of force are at an arbitrary angle to the direction of gravity. A detailed investigation of the influence of the gravitational field on hydromagnetic wave propagation is carried out in the case where the sound wave velocity is small compared to the Alfvén wave velocity, and this is used as an introduction to the general case.

The equations of hydromagnetics are continuum equations and imply that the constituent particles collectively interact. The limiting conditions in the exosphere are considered in Part 4. The generation of macroscopic hydromagnetic waves depends on whether the disturbances have a length greater than some characteristic length of the medium. This length is the Larmor radius of the positive ions for the propagation of V and T waves. If the disturbance has a characteristic length short compared with the Larmor radius but small compared with the mean free path, then collisionless hydromagnetic waves are propagated. The critical length involved in the propagation of the P mode is the distance that a particle must travel before exchanging an appreciable fraction of its momentum. Because of the long-range nature of the Coulomb forces, the particle may traverse distances short compared with the free path and still exchange a substantial proportion of its momentum. For representative models of the exosphere, disturbances at several earth radii must have a length scale large compared with several kilometers in order for acoustic waves to be propagated. V and T waves are propagated if the length scale is large compared with some hundreds of meters in the outer regions of the exosphere.

The characterization of sources is taken up in Part 5. The V mode is represented by a line in the diagnostic diagram, the T mode by a region. The V mode should be characterized by a narrow spectrum, the T mode by a much broader spectrum. The V mode is generated only if the source produces vorticity. Examples of sources generating vorticity include the outgassing of a rocket motor, the wake of a satellite in the lower exosphere, and the breakup of a finite amplitude wave. The P and T modes are generated by pressure pulses. A typical source might be the influx of ionizing radiation into the upper ionosphere giving a pressure pulse of charged particle gas. Such a pressure pulse would generate primarily spherically spreading T waves and one-dimensional acoustic waves, but if the discontinuity were sharp enough vorticity might result. The proportion of T and P waves that are generated depends on a comparison of the characteristic times and length scales of the source with the regularity and wave length of allowed waves. In the exosphere, waves of a given frequency will be predominantly P waves if the source has a small-length scale (meters), and T waves if the source has a larger-length scale (kilometers). A satellite traveling in the lower exosphere will generate P waves and possibly V waves, but it will not excite the T mode.

Part 6 treats the free oscillations of the separable V mode. The oscillations are analogous to the oscillations of a weighted spring, and have a period dependent on the geomagnetic latitude of the line of force. The free periods of these oscillations are comparable to the periods observed in the low-frequency micropulsations. The pearl-type micropulsation is interpreted as the excitation of a V mode with a narrow spectrum. The time representation of this narrow band is then a modulated sine wave characteristic of the pearl micropulsation. The propagation of acoustic waves down the lines of force is analogous to the propagation of sound waves in an ear trumpet because of convergence of the lines of force. As a result of this convergence, sound waves of finite amplitude should develop along the lines of force. The high viscosity in the outer exosphere controls the diffusion of the wave in this region.

1. SLOWNESS AND PHASE SURFACES

The propagation of magnetohydrodynamic

waves is anisotropic. As will be shown, even in the absence of dissipation, waves are dispersive at low frequencies because of the action of gravity. The description of propagation in an anisotropic medium is aided by the construction of surfaces (slowness surfaces) in wave-number space at a fixed frequency and the inversion of these surfaces to form surfaces of constant phase. In this section we review certain of the geometrical features of the slowness surface and the surface of constant phase.

The propagation of a wave can be described in terms of a phase function $\varphi(x_i, t)$. The surface

$$\varphi(x_i, t) = \text{constant}$$

at a given time might correspond to a wave crest or trough. The phase velocity, C_i , is the velocity of the surface. The wave number, k_i , is defined as the normal to the phase surface

$$k_i = \frac{\partial \varphi}{\partial x_i}$$

and is in the direction of the phase velocity. The magnitude of the phase velocity is

$$|C_i| = -\frac{\partial \varphi}{\partial t} / \left| \frac{\partial \varphi}{\partial x_i} \right| = \frac{\omega}{|k_i|}$$

where the frequency ω is defined by

$$\omega = -\frac{\partial \varphi}{\partial t}$$

The wave front is at right angles to the vector k_i , and the wave length is $2\pi/|k_i|$.

The equations for a given problem define a dispersion relation, a relation between k_i , x_i , and ω . For convenience, the dispersion relation is written

$$W(\omega, k_i, x_i) = 0$$

The group velocity

$$U_i = \frac{dx_i}{dt}$$

is determined from

$$U_i = \left(\frac{\partial W}{\partial k_i} \right) \left(\frac{\partial W}{\partial \omega} \right)^{-1}$$

The existence of a scalar function, φ , corresponds to the conservation of the number of waves, or equivalent of k_i . The existence of φ implies

$$\frac{\partial^2 \varphi}{\partial x_i \partial t} - \frac{\partial^2 \varphi}{\partial t \partial x_i} = 0$$

$$\frac{\partial k_i}{\partial t} + \frac{\partial \omega}{\partial x_i} = 0 \quad (3)$$

Using equations 2 and 3, the equation of variation of wave number takes the form

$$\frac{\partial k_i}{\partial t} + U_j \frac{\partial k_i}{\partial x_j} = - \left(\frac{\partial W}{\partial \omega} \right)^{-1} \frac{\partial W}{\partial x_i} \quad (4)$$

We used the fact that k_i is an irrotational vector (eq. 1). Equation 4 can be regarded as a form of Hamilton's equation [Eckart, 1948; Synge, 1957]. The left-hand side is the rate of change of the wave number following a point moving with velocity U_j . U_j is the velocity of a group of waves having the wave number k_i . The energy of waves is conserved, provided the dispersion relation is independent of the spatial coordinates; the right-hand side of equation 4 is not. The coefficients of the partial differential equation describing the general physical situation will be a function of position, and the spatial and time sides will be finite. The wave number and frequency must then be interpreted as the wave propagation vector and frequency. The quantities can be determined by local measurements, and their propagation in space is fixed by equation 4. Once k_i is determined, ω can be determined by the dispersion relations and equation 1. It is possible to determine phase function φ from equation 1.

The above approach yields a rigorous description of wave propagation, provided that a phase function exists. The usual treatment of propagation of waves in an inhomogeneous medium involves two kinds of approximations. The field function can be written as the product of an amplitude function F , and a phase function G .

$$V = F(x_i, t) G[\varphi(x_i, t)]$$

The variation of the field in time and space is

$$\frac{\partial V}{\partial t} = \frac{\partial F}{\partial t} G - F \frac{\partial G}{\partial \varphi} \omega$$

$$\frac{\partial V}{\partial x_i} = \frac{\partial F}{\partial x_i} G + k_i F \frac{\partial G}{\partial \varphi}$$

The W.K.B. approximation neglects the first term relative to the second term on the right-

hand side. This implies that the typical length scale associated with the spatial inhomogeneity is large compared with the wave length. For an infinite medium the phase function G can be taken as $G = \exp(i\varphi)$ with the result that the operations of differentiation with respect to time and space yield

$$\frac{\partial V}{\partial t} = -i\omega V; \quad \frac{\partial V}{\partial x_i} = ik_i V$$

Under a similar but not identical approximation, the phase function is assumed linear in k_i and ω . Let L be the length scale in a given direction. Then for times limited by

$$Ut \ll L$$

the phase function is approximately

$$\varphi = k_i x_i - \omega t$$

In the problems treated below, the dispersion relation will be a function of a single coordinate, say x_3 , and then the phase function will be of the form

$$\varphi = k_1 x_1 + k_2 x_2 + \int k_3 dx_3 - \omega t$$

At a given frequency and point in space it is possible to construct a three-dimensional surface in wave number space determined by

$$W(\omega, k_i, x_i) = 0$$

This surface will be termed the slowness surface [Synge, 1957]. The slowness surface and the surface of constant phase are connected by $k_i x_i - \omega t = 0$. Every point on the slowness surface corresponds to a wave. Some of the waves are disallowed because of the radiation conditions. The radiation condition allows only the propagation of energy outward from the source in an initial-value problem.

The surface of constant phase can be constructed from the slowness surface. The surface of constant phase is the envelope at a given time of waves originating from a source at time $t = 0$. The direction of the wave resulting from a point on the slowness surface is along the normal to the slowness surface. It should be noted that the direction of the wave number, k_i , and the normal to the slowness surface need not be the same. The crest and troughs of the waves are then not necessarily at right angles to the di-

rection of propagation. Since the normal to the slowness surface is parallel to the direction x_i , we have

$$x_i = \text{constant} \frac{\partial W}{\partial k_i} \quad (5)$$

where at unit time

$$k_i x_i = \omega \quad (6)$$

Combining equations 5 and 6, the surface of constant phase at unit time is given by

$$x_i = - \left(\frac{\partial W}{\partial k_i} \right) \left(\frac{\partial W}{\partial \omega} \right)^{-1} \quad (7)$$

At unit time the surface of constant phase is, then, a representation of the group velocity. The above interpretation runs into difficulty when the group velocity is a function of the spatial coordinate. In this case one can only calculate the initial wave form. The wave form is distorted as the wave progresses through the medium. A further complication arises from dispersion. Waves are dispersed if $W(k_i, x_i, \omega)$ is not homogeneous in k_i and ω . The phase surface for different frequencies will spread in time. A pulse contains a range of frequencies, and there will be a range of phase velocities for a fixed direction.

Once the slowness surface is determined the surface of constant phase can be calculated using equation 7. A geometric interpretation of equation 7 is that the vector x_i is in the direction of a perpendicular dropped from the tangent plane of the slowness surface to the origin. The magnitude of the vector x_i is proportional to the frequency and inversely proportional to the length of the perpendicular passing through the origin. x_i is the pole, with respect to a sphere of radius $\omega^{1/2}$ with a center at the origin, of the tangent plane to the slowness surface. The pole of the tangent plane is then the reciprocal pole of the slowness surface with respect to the sphere of the radius $\omega^{1/2}$. The description of the wave-number surface as a slowness surface results from the fact that the phase surface represents the progress in time of the wave front, and its reciprocal is the surface in wave number space.

The geometrical characteristics of the slowness surface determine the geometrical configuration of the surface of the constant phase. A qualitative picture of surface of constant phase can

be obtained from the slowness surface without detailed calculations if certain geometrical correspondences are noted. Each tangent plane to the slowness surface corresponds to a point on the surface of constant phase. Reciprocally, to every tangent plane on the surface of constant phase there corresponds a point on the slowness surface. A tangent plane at a point of the slowness surface with N contacts corresponds to a point on the wave surface with N tangent planes. A parabolic point (a point at which one or both of the principal curvatures vanishes [Weatherburn, 1955; Eisenhart, 1947]) on the slowness surface is reciprocally a point on an edge of regression on the surface of constant phase. Further, if a tangent plane to the slowness surface touches the surface on a curve of tangency, all points on the curve correspond to a single point, conical point, on the surface of constant phase. As an illustration of these geometrical correspondences, consider a plane-slowness surface with a circular bump on it. In the region away from the bump the tangent plane to the slowness surface is in contact at an innumerable number of points. The tangent plane corresponds to a single point on the surface of constant phase. Associated with the bump will be a circle of parabolic points. This circle maps into a circular cusp in the surface of constant phase surface. As the bump is reduced and approaches a plane, the surface of constant phase approaches a point. If the slowness surface is a plane, then the propagation is one-dimensional.

A picture of the relative amplitudes of waves traveling in different directions can also be obtained from a consideration of the slowness surface. Each tangent plane to the slowness surface corresponds to a wave. A tangent plane with many points of contact represents the sum of these waves. The amplitude of that part of the phase surface corresponding to the tangent plane with many points of contact should be greater than the amplitude of that part of the phase surface resulting from a tangent plane with a single point of contact to the slowness surface. As an example, the conical point on the surface of constant phase corresponding to a curve of tangency on the slowness surface should have a larger amplitude after a given distance of propagation than the corresponding points on the constant phase surface with

angle tangent point. If the slowness surface is a sphere, then the corresponding surface of constant phase is also a sphere, and the amplitude decays as $1/r$. If, on the other hand, the surface is a plane, there is no geometrical attenuation. These two cases provide limits for the geometrical attenuation of waves. *Lighthill* [1960], in a detailed investigation using the saddle-point method in the asymptotic evaluations, finds that points on cuspidal edge decay as $r^{-5/6}$ while conical points decay as $r^{-1/2}$.

The determinations of the surface of constant phase can be calculated as above for dispersive waves, but the spread in time must be considered. If the surface of constant phase is a sphere, the disturbance grows as a hollow spherical shell and the geometric attenuation is $r^{-3/2}$.

The geometrical correspondences between the slowness surface and the surface of constant phase noted above will be illustrated in detail in the treatment of hydromagnetic waves perturbed by gravity.

2. PARALLEL MAGNETIC AND GRAVITATIONAL FIELDS

Equations of motion. The simplest description of the motion of an electrically conducting fluid in the presence of a gravitational field with potential gz is

$$\frac{dv_i}{dt} = -g \frac{\partial z}{\partial x_i} - \frac{1}{\rho} \frac{\partial p}{\partial x_i} - \frac{1}{4\pi\mu} \left[B_j \frac{\partial B_i}{\partial x_j} - B_i \frac{\partial B_j}{\partial x_i} \right] \quad (8)$$

$$\frac{dB_i}{dt} = B_j \frac{\partial v_i}{\partial x_j} - B_i \frac{\partial v_j}{\partial x_i} \quad (9)$$

$$\frac{d\rho}{dt} + \rho \frac{\partial v_i}{\partial x_i} = 0 \quad (10)$$

$$\frac{ds}{dt} = \frac{q}{T} \quad (11)$$

where the gravitational constant at some reference altitude is g and the gravitational altitude is $z(x_i)$. A source term q , the net accession of heat, is included in the equation of entropy conservation (eq. 11).

The description yields a good approximation, provided that a number of restrictions are met:

1. There exists a scalar fluid pressure, p , which is a function of the entropy, s , and density, ρ .

2. Ohm's law is taken in the form $E_i = \epsilon_{ijk} B_j v_k$.

3. The usual assumptions of magnetohydrodynamics are made in that the displacement current is neglected and all velocities are small compared with the velocity of light.

4. Dissipation due to the finite viscosity and thermal conductivity of the fluid is neglected. In addition, Coriolis forces are neglected, although the effects of rotation on the gravitational field can be included by appropriate choice of the gravitational altitude.

The equation of the conservation of mass (eq. 10) can be combined with the thermodynamic relations to relate pressure and density variations.

$$\frac{dp}{dt} - C^2 \frac{d\rho}{dt} = \frac{\rho(\gamma - 1)}{\alpha} \frac{q}{T} \quad (12)$$

C is the velocity of sound while γ is the ratio of the specific heats, c_p/c_v .

The set of nonlinear equations 8-12 can be reduced to a tractable linear form by the perturbation scheme

$$\rho = \rho_0 + \rho_1 + \dots \quad q = q_1 + \dots$$

$$s = s_0 + s_1 \quad B_i = B_i^0 + B_i^1 + \dots$$

$$p = p_0 + p_1 + \dots \quad v_i = v_1 + \dots$$

where the quantities p_0 , B_i^0 , etc., represent the initial state of the fluid and field. Maxwell's equations (9) are satisfied identically to zero order.

Equation 1 yields

$$\frac{\partial p_0}{\partial x_i} = -\rho_0 g \frac{\partial z}{\partial x_i} - \frac{1}{4\pi\mu} \left[\frac{1}{2} \frac{\partial B_{ij}^2}{\partial x_i} - B_i^0 \frac{\partial B_i^0}{\partial x_j} \right] \quad (13)$$

The relative magnitude of the two terms on the right-hand side of equation 13 determines whether the initial particle distribution is gravitationally or magnetically controlled. For the purposes of the present paper we will make the usual assumption that gravity determines the particle distribution over most of the exosphere. This need not hold true, and the particle distribution may be influenced greatly by the magnetic field.

The equations to first order are then

$$\frac{\partial B_i}{\partial t} = B_i^0 \frac{\partial v_i}{\partial x_i} + B_i^0 \frac{\partial v_i}{\partial x_i} - v_i \frac{\partial B_i^0}{\partial x_i} \quad (14)$$

$$\rho_0 \frac{\partial v_i}{\partial t} = -\rho_0 g \frac{dz}{\partial x_i} - \frac{\partial p}{\partial x_i} - \frac{1}{4\pi\mu}$$

$$\left[B_i \frac{\partial B_i^0}{\partial x_i} + B_i^0 \frac{\partial B_i}{\partial x_i} - B_i^0 \frac{\partial B_i}{\partial x_i} - B_i \frac{\partial B_i^0}{\partial x_i} \right] \quad (15)$$

$$\frac{\partial p}{\partial t} + v_i \frac{\partial p_0}{\partial x_i} = -\rho_0 C_0^2 \frac{\partial v_i}{\partial x_i} \quad (16)$$

$$\frac{\partial s}{\partial t} + \frac{\partial s_0}{\partial x_i} v_i = \frac{q_i}{T_0} \quad (17)$$

where the unprimed variables represent the perturbed quantities. The initial pressure and entropy gradients appear in the first order equations. These quantities are not independent. Using the standard thermodynamic relations

$$\frac{\partial T_0}{\partial x_i} = \frac{(\gamma_0 - 1)}{\alpha_0} \frac{\partial \rho_0}{\partial x_i} + \frac{T_0}{c_v^0} \frac{\partial s_0}{\partial x_i};$$

$$\frac{\partial p_0}{\partial x_i} = C_0^2 \frac{\partial \rho_0}{\partial x_i} + \frac{(\gamma_0 - 1)}{\alpha_0} \rho_0 \frac{\partial s_0}{\partial x_i}$$

we obtain for the entropy gradient

$$\frac{\partial s_0}{\partial x_i} = \frac{c_p^0}{T_0} \left(\frac{\partial T_0}{\partial x_i} + \frac{g \alpha_0 T_0}{c_p^0} \frac{\partial z}{\partial x_i} \right)$$

where the adiabatic temperature gradient is

$$\frac{\partial T_A}{\partial x_i} = \frac{g \alpha_0 T_0}{c_p^0} \frac{\partial z}{\partial x_i}$$

α_0 is the coefficient of thermal expansion in the initial state. The velocity of sound, coefficient of thermal expansion and heat capacity at constant pressure satisfy the simple relations

$$C_0^2 = \gamma_0 R T_0 \quad c_v^0 = R/(\gamma_0 - 1) \quad \alpha_0 = T_0^{-1}$$

in an ideal gas. An ionized gas composed equally of electrons and singly ionized atoms has an effective value, $\gamma_0 = 4/3$, since the value of γ_0 for the ion is $5/3$ and the electrons will behave more or less isothermally corresponding to $\gamma_0 = 1$.

Parallel magnetic and gravitational fields. A marked simplification results in the case of a constant magnetic field with lines of force parallel to the gravitational field. The fields and mo-

tions are cylindrically symmetrical. The initial magnetic field is taken along the x_3 axis and the effect of curvature of the surfaces of equipotential are neglected.

$$B_i^0 = (0, 0, B^0); \quad \frac{\partial z}{\partial x_i} = \left(0, 0, \frac{\partial z}{\partial x_3} \right)$$

The first order perturbation equations reduce

$$\frac{\partial b_1}{\partial t} = \frac{\partial v_1}{\partial x_3} \quad \frac{\partial b_2}{\partial t} = \frac{\partial v_2}{\partial x_3}$$

$$\frac{\partial b_3}{\partial t} = -\left(\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} \right) \quad (1)$$

$$\frac{\partial v_1}{\partial t} + \frac{1}{\rho_0} \frac{\partial p}{\partial x_i} + C_A^2 \left[\frac{\partial b_3}{\partial x_1} - \frac{\partial b_1}{\partial x_3} \right] = 0$$

$$\frac{\partial v_2}{\partial t} + \frac{1}{\rho_0} \frac{\partial p}{\partial x_2} + C_A^2 \left[\frac{\partial b_3}{\partial x_2} - \frac{\partial b_2}{\partial x_3} \right] = 0 \quad (1)$$

$$\frac{\partial v_3}{\partial t} + \frac{1}{\rho_0} \frac{\partial p}{\partial x_3} + \frac{1}{\rho_0} \frac{g}{C_0^2} p - N^2 Q = 0$$

$$\frac{\partial p}{\partial t} - g \rho_0 v_3 + \rho_0 C_0^2 \left(\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3} \right) = 0 \quad (1)$$

$$\frac{\partial Q}{\partial t} + v_3 = 0 \quad (1)$$

b_i is the normalized fluctuation in the magnetic field.

$$b_i = B_i/|B_i^0|$$

C_A is the Alfvén wave velocity in the undisturbed medium.

$$C_A = \frac{|B_i^0|}{(4\pi\mu\rho_0)^{1/2}}$$

Q is the normalized fluctuation in entropy.

$$Q = S \left(\frac{ds_0}{dx_i} \right)^{-1} \frac{\partial z}{\partial x_i}$$

N is the Väisälä frequency [Eckart, 1960].

$$N^2 = g \frac{\partial s_0}{\partial x_i} (\gamma_0 - 1) / \left(\alpha_0 C_0^2 \frac{\partial z}{\partial x_i} \right)$$

We interpret N as the frequency of oscillations of a small mass of fluid when displaced from its zero-order position in the gravitational field and allowed to move freely. As will be shown, motions with frequencies large compared

not influenced by gravity. In a perfect gas the Väisälä frequency takes the form

$$N^2 = g \left[\frac{dT_0}{dz} + \frac{g(\gamma - 1)}{\gamma R} \right] T_0^{-1}$$

if the fluid is isothermal then

$$N^2 = \frac{g^2}{C_0^2} (\gamma_0 - 1)$$

Equations 14a, 15a, 16a, and 17a form a coupled set for v_i . The complexity is reduced in a homogeneous fluid by choosing a new set of variables. More importantly, one mode of motion is separable [Grad, 1959; Kanwal and Reddell, 1960; Carstou, 1960]. A similar separation is possible in the presence of gravitational fields provided gravity and the lines of force are parallel. We consider a motion in which $v_0 = 0$; $v_s = 0$; $b_s = 0$. The pressure equation implies

$$\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} = 0$$

where there is no variation of entropy.

$$\frac{\partial Q}{\partial t} = 0 \quad Q = 0$$

Equations 14a and 15a combine to give

$$\frac{\partial \xi}{\partial t} = C_A^2 \frac{\partial j}{\partial x_3} \quad \frac{\partial j}{\partial t} = \frac{\partial \xi}{\partial x_3} \quad (18)$$

where ξ is the vorticity, or spin, of the fluid about the line of force and j is the normalized current density.

$$\xi = \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \quad j = \frac{\partial b_2}{\partial x_1} - \frac{\partial b_1}{\partial x_2}$$

Equations 18 yield

$$\frac{\partial^2 \xi}{\partial t^2} = C_A^2 \frac{\partial^2 j}{\partial x_3^2} \quad \frac{\partial^2 j}{\partial t^2} = \frac{\partial}{\partial x_3} \left(C_A^2 \frac{\partial j}{\partial x_3} \right)$$

where the vorticity about the line of force is propagated along the lines of force at the Alfvén wave velocity as a one-dimensional wave. The thermodynamic state of the system remains constant throughout the variations in entropy and pressure. The wave velocity varies with varying density of the fluid.

The P and T modes are not separable. The separation of pressure, longitudinal component

of the magnetic field, component of velocity along the magnetic field, and entropy are governed by the system of equations

$$\frac{\partial^2 b_3}{\partial t^2} - \frac{1}{\rho_0} \left(\frac{\partial^2 p}{\partial x_1^2} + \frac{\partial^2 p}{\partial x_2^2} \right) - C_A^2 V^2 b_3 = 0$$

$$\frac{\partial v_3}{\partial t} + \frac{1}{\rho_0} \left(\frac{\partial}{\partial x_3} + \frac{g}{C_0^2} \right) p - N^2 Q = 0$$

$$\frac{\partial p}{\partial t} + \rho_0 C_0^2 \left(\frac{\partial}{\partial x_3} - \frac{g}{C_0^2} \right) v_3 - \rho_0 C_0^2 \frac{\partial b_3}{\partial t} = 0$$

$$\frac{\partial Q}{\partial t} + v_3 = 0 \quad (19)$$

We note that equation 19 reduces to a spherically spreading Alfvén wave for $p = 0$, $g = 0$.

As a preliminary to the consideration of wave propagation in an inhomogeneous medium, we consider equation 19 for $g = 0$. The dispersion relation is

$$\kappa^2 = \frac{(\omega^2 - C_0^2 k_3^2)(\omega^2 - C_A^2 k_3^2)}{(C_A^2 + C_0^2)\omega^2 - C_A^2 C_0^2 k_3^2} \quad (20)$$

where κ is

$$\kappa^2 = k_1^2 + k_2^2 \quad (21)$$

The propagation is thus symmetrical about the x_3 axis.

Given values of the wave number k_3 and frequency ω equation 21 determines κ^2 . The value thus determined may be negative. However, κ must be real, since the system is conservative and is assumed stable. The values of frequency and wave number resulting in a negative κ do not give an allowed eigen solution. The regions in the $\omega - k_3$ plane of permitted waves are shown in Figure 1. The upper left-hand region of the diagram is bounded by the line $\omega = C_A k_3$. Over this region $\kappa^2 > 1$. A second region of allowed solutions is bounded by the lines $\omega = C_0 k_3$ and

$$\omega = \left(\frac{C_A - C_0}{C_0^2 + C_A^2} \right)^{1/2}$$

Along this last line, the wave number κ becomes infinite. This is interpreted physically as a vanishing wave length.

The diagram is scaled assuming

$$\beta = C_0/C_A = 0.5$$

The upper left-hand region corresponds to the propagation of the T mode in which two-dimen-

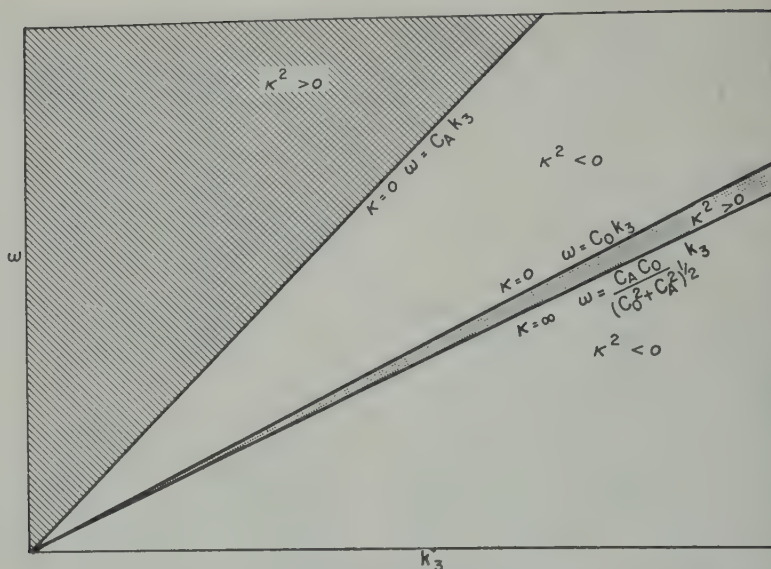


Fig. 1. Diagnostic diagram for hydromagnetic wave propagation in a uniform fluid. Diagram is scaled so that $C_0/C_A = 0.5$. The upper region is a region of T mode propagation. The narrow stippled region is the P mode region.

sional divergence is propagated at speeds comparable to the Alfvén wave velocity. The narrow region corresponds to the propagation of the P mode. The lines bounding the regions of allowed solutions are straight, since the dispersion relation is homogeneous in the frequency and wave number; the T and P modes are propagated in a uniform medium without dispersion. It should be noted that the line $\omega = C_A k_3$ corresponds to the V mode, which has been separated by the choice of variables.

The dispersion relation 20 may be written in the form

$$W(k_i, \omega) = \beta^2 k_3^4 + [\beta^2 \kappa^2 - (\beta^2 + 1) l^2] k_3^2 + l^2(l^2 - (1 + \beta^2) \kappa^2) = 0 \quad (22)$$

where $\beta = C_0/C_A$, $l = \omega/C_A$ so that the unit of distance is C_A/ω . At a fixed frequency ω the slowness surface corresponding to the diagnostic diagram Figure 1 is shown in Figure 2. The complete surfaces are obtained by rotating the $k_3 - \kappa$ plane about the k_3 axis and reflecting across the plane $k_3 = 0$. The two sheets of the slowness surface determined by the dispersion relation 22 are a sphere flattened along the κ direction and a near plane with a bump towards the origin. In addition, the plane representing the one-dimensional propagation of the vorticity

(V mode) has been added. The near sphere corresponds to the spherically spreading mode. This is noted by allowing β to approach zero in equation 22. As β approaches zero, propagation surface becomes a sphere corresponding to the spherical propagation of Alfvén waves.

Substituting $k_3 = l$ into the pressure equation 19 we obtain

$$\frac{p}{p_m} = \frac{\beta^2}{1 - \beta^2}$$

where p_m is the variation in the magnetic pressure.

$$p_m = \rho_0 C_A^2 b_3 = \frac{B^0 B_3^1}{4\pi\mu}$$

We note that in the T mode a small fluctuation in particle pressure results in a larger fluctuation in the magnetic perturbation pressure, provided $\beta < 1$. The T mode propagates a fluctuation in the longitudinal magnetic field plus a condensation of the charged particles.

Over most of the planar sheet we

$$k_3 = \omega [C_0^{-2} + C_A^{-2}]^{1/2}$$

which corresponds to the one-dimensional propagation at a modified sound velocity.

$$C \text{ phase} \cong \frac{\omega}{k_3} = \frac{C_A C_0}{(C_0^2 + C_A^2)^{1/2}}$$

er to obtain the variation of pressure in mode we substitute (23) into the pressure on, with the result

$$p + p_m = 0$$

the excess pressure, p , and the magnetic re, p_m , are in approximate balance in the de. As p drops and suction develops, a etic field is generated. As the particle re increases, energy is transferred from agnetic to the acoustic field.

T mode with a nearly spherical slowness e will undergo geometrical attenuation as The V mode with a planar slowness surface goes no geometrical attenuation. The gation of the P mode is more complicated.

reading of the P mode. The surfaces of ant phase can be constructed from the ess surfaces. The wave form at unit time wn in Figure 3 for a source located at the at time $t = 0$ and for $\beta = 0.6, 0.8$. The etic lines of force are taken along the x_3 and there is cylindrical symmetry about x_3 axis, and across the $x_2 - x_3$ plane. The of distance is $C_A t$. The slowness surfaces for V mode degenerate to a point on the x_3 axis. T mode, represented by a sphere flattened the κ direction in wave number space, into a sphere flattened in the x_3 direction. inverse property of the slowness surface is rated in the wave surface for the T mode. phase velocity in the direction perpendicu- to the lines of force is greater than the ity along the lines of force.

one-dimensional P mode has a constant e surface in the form of a cusped surface a conical point on the x_3 axis. The cusped ce is subtended by a cone with a vertex at origin. The angular diameter of the cone ases with increasing β up to $\beta = 1$.

propagation of the P mode (for $\beta < 1$) some ways analogous to the phenomenon of al refraction in optics of anisotropic crys- [Bazer and Fleischman, 1959]. The slowness e has a dimple which is concave in the ion of the magnetic field. The presence of imple leads to a circle of parabolic points, t is these parabolic points that give rise to uspidal edge on the phase surface.

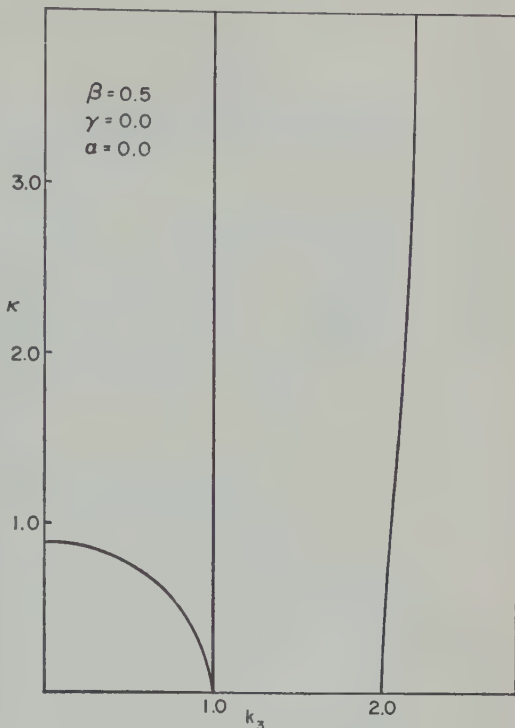


Fig. 2. The slowness surfaces for the propagation of hydromagnetic waves in a uniform medium. The complete surfaces are obtained by rotation about the k_3 axis and reflection across the κ plane. The unit of wave number is ω/C_A . The lines of force are parallel to the k_3 direction. The near sphere corresponds to a T mode propagation; the plane at $k_3 = \pm 1$ is the V mode propagation and the plane with a bump describes the P mode propagation.

For large values of κ , the slowness surface becomes tangent to a plane perpendicular to the x_3 axis. The plane corresponds to the conical point on a surface of constant phase. The points along the cuspidal edge geometrically attenuate at a slower rate than all other points on the surfaces excepting the conical point, which decays at the slowest rate. The P mode is then dominated by the cuspidal circle and conical point. A projection of the wave amplitude onto a plane perpendicular to the direction of propagation would show a circle around a central dot.

The transition from $\beta < 1$ to $\beta > 1$ is illustrated in Figure 4. As the sound velocity is increased at constant Alfvén wave velocity, the phase surface for the P mode becomes a flattened sphere and the phase surface for the T mode is a cusped surface bounded by a cone

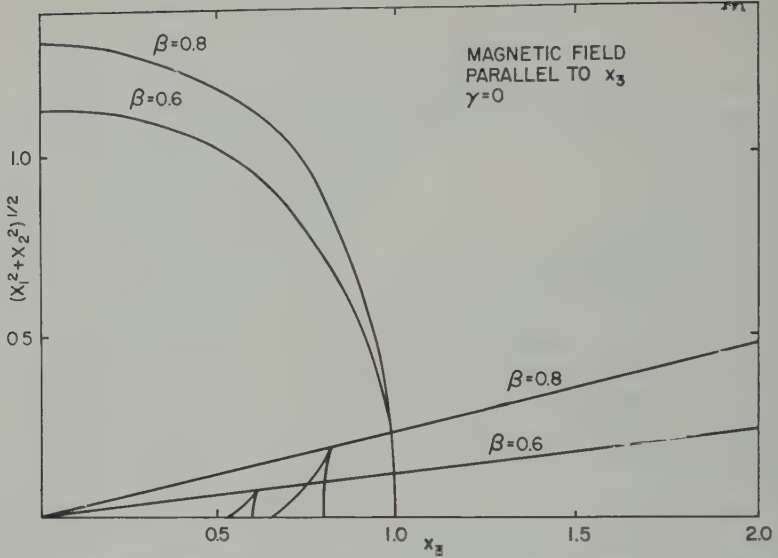


Fig. 3. Surfaces of constant phase for a uniform medium. The complete surfaces are obtained by rotation about the x_3 axis and reflection across the $x_1 - x_2$ plane. The cusped triangles correspond to a P wave, the flattened sphere to a T wave.

with a vertex at the origin. The angular diameter of the cone bounding the T mode becomes less with increasing β .

Over much of the exosphere β is very small. In the limit of small β , the cusped triangle generating the P phase surface reduces to a point maintaining the cuspidal form. The propagation of the P mode is then very nearly one dimensional.

The T mode propagation becomes nearly spherical as β decreases. At high frequencies, energy may be propagated in the exosphere by one-dimensional V waves, a near one-dimensional P wave, and a spherical spreading T wave.

Gravitational coupling. The field equations 14a-17a combine to yield the dispersion relation

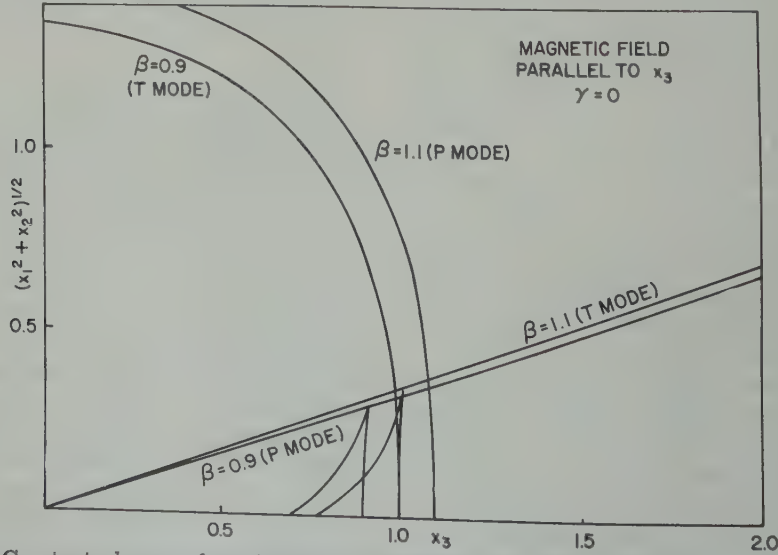


Fig. 4. Constant phase surfaces illustrating transitions from T mode propagation to P mode propagation.

$$\kappa^2 = \frac{[(\omega^2 - N_1^2) - C_0^2 k_3^2][\omega^2 - C_A^2 k_3^2]}{(C_0^2 + C_A^2)\omega^2 - C_A^2 N_1^2 - C_0^2 N^2 - C_A^2 C_0^2 k_3^2} \quad (24)$$

modified Väisälä frequency.

$$N_1^2 = N^2 + \frac{g^2}{C_0^2}$$

It should be noted that the dispersion relation applies only to the coupled P and T modes. The T mode is unaffected by gravity, provided (x, t) is along the line of force.

The diagnostic diagram corresponding to equation 24 is shown in Figure 5. The regions of allowed propagation are limited by the curves of vanishing wave number κ

$$\omega = C_A k_3$$

$$\omega^2 = N_1^2 + C_0^2 k_3^2$$

and the curve

$$(C_0^2 + C_A^2)\omega^2 - C_0^2 C_A^2 k_3^2$$

$$- C_0^2 N^2 - C_A^2 N_1^2 = 0$$

which κ becomes infinite, corresponding to zero wave length. The curves bounding regions of allowed propagation are no longer straight lines, the dispersion relation is not homogeneous in the wave number and frequency.

For frequencies near N_1 , the propagation is dispersive.

At large frequencies and wave number, k_3 , the diagnostic diagram is similar to that shown in Figure 1. There are two modes of propagation; a P mode and a T mode. Gravity dominates the propagation at low frequencies and large wave numbers. The characteristic frequencies noted in Figure 5 are N_1 ,

$$\omega_1 = N_1/(1 - \beta^2)^{1/2}$$

$$\gamma_1 = N_1/\omega_1 = (1 - \beta^2)^{1/2}$$

$$\omega_2^2 = N_1^2 + \beta^2 N^2$$

$$\gamma_2^2 = \frac{N_1^2}{\omega_2^2} = \left(1 + \beta^2 \frac{N^2}{N_1^2}\right)^{-1}$$

$$N_2^2 = \frac{C_0^2 N^2 + C_A^2 N_1^2}{C_0^2 + C_A^2} = \frac{N_1^2 + \beta^2 N^2}{1 + \beta^2}$$

For $\beta < 1$, the relative magnitudes of the four frequencies are $N_2 < N_1 < \omega_2 < \omega_1$.

The nature of the waves in the various regions of allowed propagation in the diagnostic diagram is best obtained by constructing the slow-

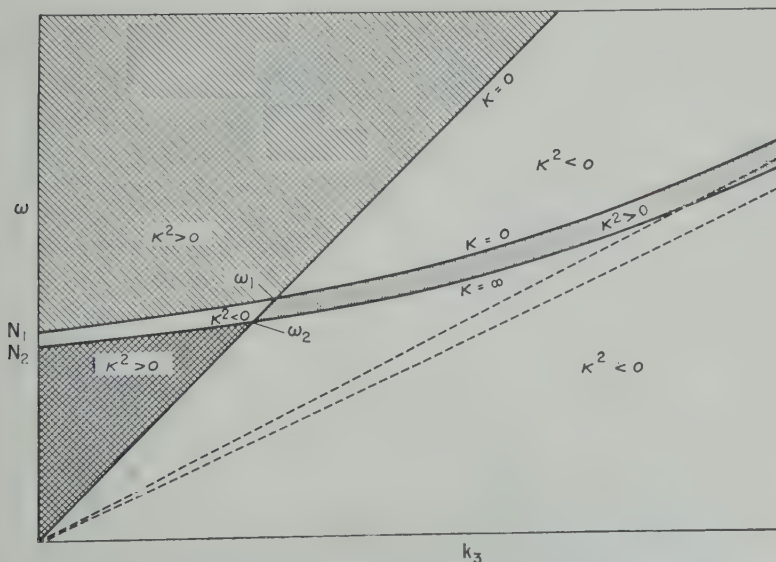


Fig. 5. Diagnostic diagram for hydromagnetic wave propagation with the lines of force parallel to the direction of gravity. Diagram scaled with $\beta = 0.5$. The upper left-hand region corresponds to perturbed T mode propagation, the narrow stippled region to P mode propagation, and the lower crosshatched region to gravitationally controlled propagation.

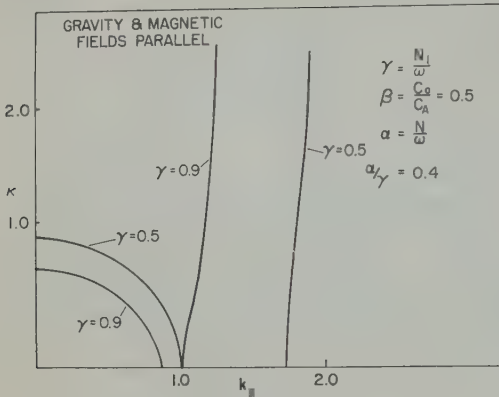


Fig. 6. Slowness surfaces for parallel gravity and magnetic fields. Complete surface is obtained by rotating about k_s direction and reflection across the $k_1 - k_s$ plane. The planar sheets correspond to perturbed P mode propagation, the distorted spheres to T mode propagation. The propagation is dispersive.

ness surface and the corresponding surfaces of constant phase. The dispersion relation (eq. 24) is written in terms of dimensionless quantities

$$\gamma = N_1/\omega \quad \alpha = N/\omega$$

as

$$\beta^2 k_s^4 - [l^2(1 - \gamma^2) + \beta^2(l^2 - \kappa^2)]k_s^2 + l^2(1 - \gamma^2)(l^2 - \kappa^2) - \beta^2 l^2(1 - \alpha^2)\kappa^2 = 0$$

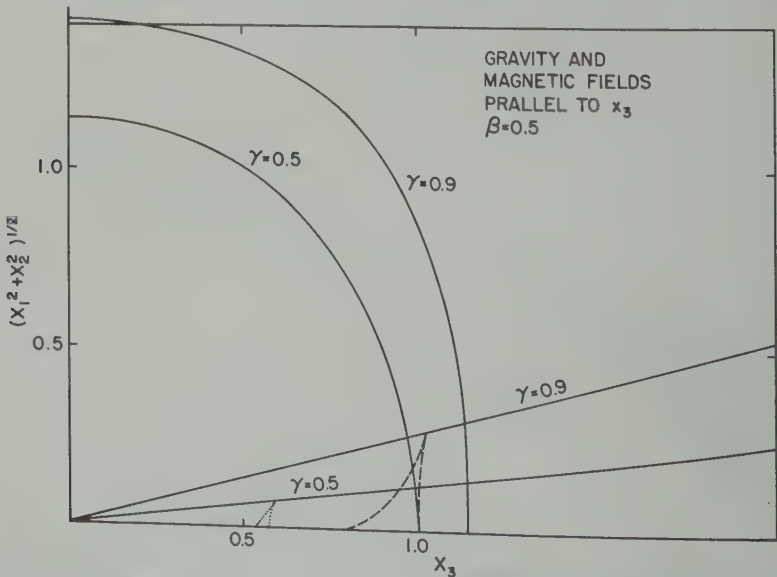


Fig. 7. The constant phase surfaces corresponding to the slowness surfaces of Figure 6. Both the T mode, nearly spherical phase surface, and the V mode, cusped surface, are highly dispersive.

The wave numbers are thus measured in units of $l = \omega/C_A$.

The slowness surface for the values of $\gamma = 0.9$ are shown in Figure 6. β is set at 0.5. The surface is symmetric about the k_s axis and across the $k_1 - k_s$ plane. The propagation of both the modified T and P modes is dispersive, with the high-frequency T and P components traveling at a slower rate than the low-frequency components. In addition to the two modes shown, the V mode is propagated without dispersion. Comparing Figure 6 with Figure 2, we note that for $\gamma = 0.5$ the T mode is affected only slightly by gravity, while the P mode is markedly changed.

The dispersion resulting from gravity is clearly illustrated in Figure 7 where the constant phase surfaces corresponding to the slowness surfaces of Figure 6 are shown. The lower-frequency wave, $\gamma = 0.9$, travels faster than the higher-frequency wave. In addition, the angular diameter of the cone enclosing the low-frequency wave is larger. This corresponds to the fact that at low frequencies in the diagnostic diagram the lines bounding the region of propagation of the P wave become more nearly parallel to the k_s axis. The values of the critical frequencies are $\gamma_1 = 0.87$; $\gamma_2 = 0.954$ for $\beta = 0.5$. At frequencies greater than $\gamma = 0.87$, the T mode is almost

turbed by the action of gravity. The transition region between frequencies ω_1 and ω_2 is illustrated in Figure 8 where the slowness surfaces for $\gamma = 0.9, 0.95$, and 1.0 are shown. As the frequency decreases with increasing γ the slow-point moves toward the origin. At frequencies less than N_1 there is no propagation in the T mode. At frequencies intermediate to ω_2 and N_1 , the propagation surface for the T mode becomes greatly flattened in the k_3 direction (see Fig. 9).

There is no propagation of the T mode for frequencies less than ω_2 where a new region appears in the diagnostic diagram. The slowness surfaces for the transition region $\omega_2 - N_2$ are shown in Figure 9. The surfaces of constant γ corresponding to the slowness surfaces are shown in Figure 10. As the frequency is lowered, the angular diameter of the cone bounding the surface of constant phase increases. It is denoted by the conical point leads the propagation surface than following, as for an unperturbed P wave. As the frequency is lowered still further,

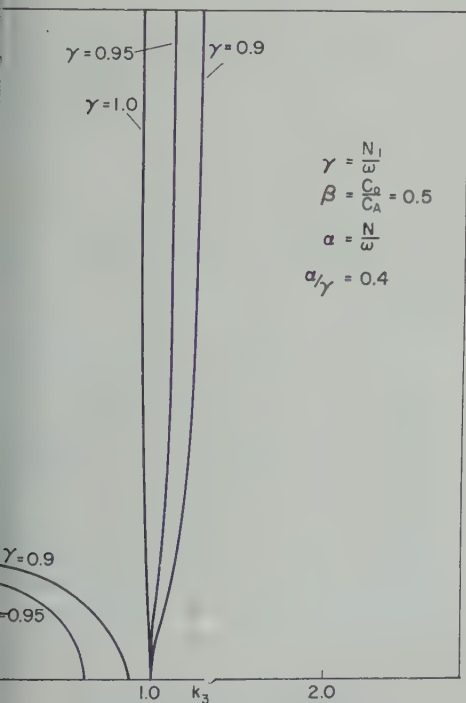


Fig. 8. Slowness surfaces at frequencies near the critical frequency N_1 . Complete surfaces are obtained by rotation about the k_3 axis and reflection across the $k_1 - k_2$ plane.

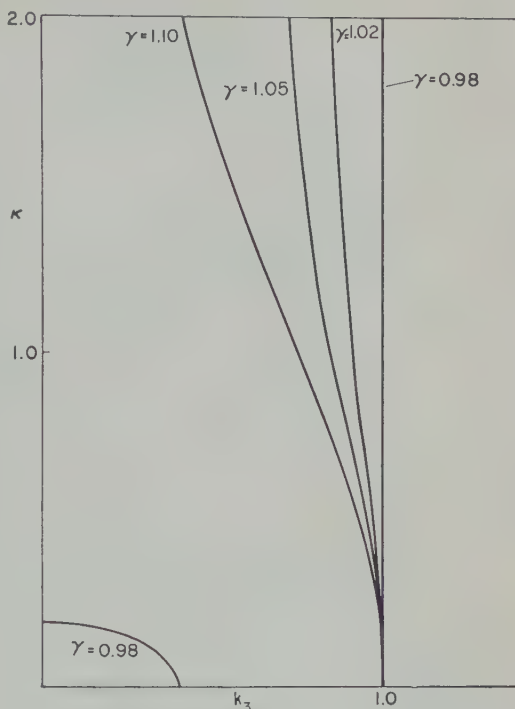


Fig. 9. Slowness surfaces at frequencies lower than N_1 .

the propagation again becomes spherical (see Figure 11 where the slowness surfaces for $\gamma = 1.1, 1.2, 1.3, 1.5$ are shown). The surfaces of constant phase corresponding to the slowness surfaces of Figure 11 are shown in Figure 12. For P propagation for $\gamma = 1.1$ is a wave spreading about the x_3 direction. For $\gamma = 1.2$ the surface becomes a near sphere but flattened in the direction perpendicular to the line of force.

Further insight into the nature of the waves in the lower left-hand corner of the diagnostic diagram is obtained by examining the relative variations of particle and magnetic pressure. Combining the equations in entropy, pressure, and v_3 , we have

$$[l^2(1 - \gamma^2) - \beta^2 k_3^2]p = (1 - \alpha^2)l^2\beta^2 p_m$$

As an example, consider propagation for the frequency corresponding to $\gamma = 1.0$ (see Fig. 8). Over most of the slowness surface $k_3 = l$ and $p + (1 - \alpha^2)p_m = 0$.

An increase in particle pressure results in a decrease in local magnetic pressure, provided $\alpha < 1$. The balance is similar to the balance that obtains in a P wave, but it is modified by

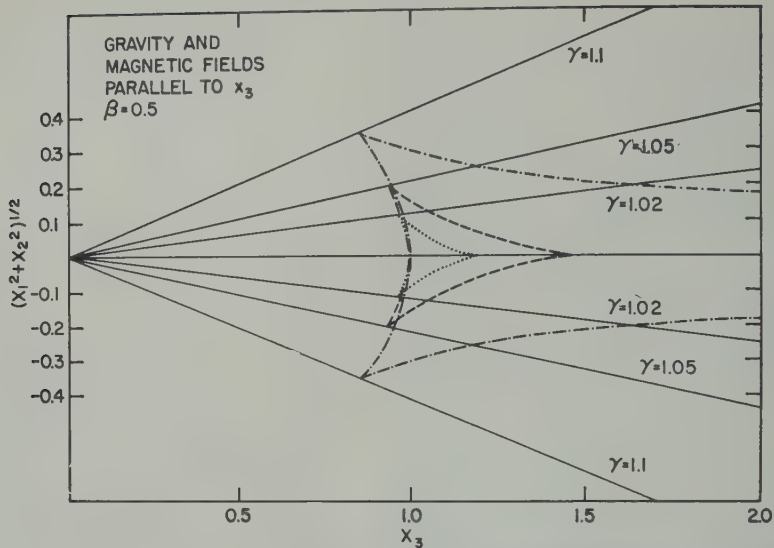


Fig. 10. Constant phase surfaces corresponding to the slowness surfaces of Figure 9. A complete surface is obtained by rotating about the line $(x_1^2 + x_2^2)^{1/2} = 0$ and across the plane $x_3 = 0$. The conical point leads the propagation.

the presence of the coefficient α multiplying the magnetic pressure. At lower frequencies the propagation approaches the propagation of a

T mode. Thus, for $\gamma = 1.5$ the propagation along the direction $k_3 = 0$ gives

$$(1 - \gamma^2)p = (1 - \alpha^2)\beta^2 p_m$$

while the propagation along the direction $\kappa = 0$ gives

$$[(1 - \gamma^2) - \beta^2]p = (1 - \alpha^2)\beta^2 p_m$$

The propagation is intermediate between a T mode and a P mode with changes in pressure being balanced in part by alterations of the magnetic pressure.

The above arguments show the importance of the frequency N_1 in separating those waves that propagate as T or P modes from those waves that propagate as modes highly modified by gravity. The frequency N is of more fundamental character from a particle point of view being the oscillation frequency of a parcel of charged particles. A further interpretation of N results from examining the phase relationship between the variations in pressure and entropy. The entropy and pressure equations combining

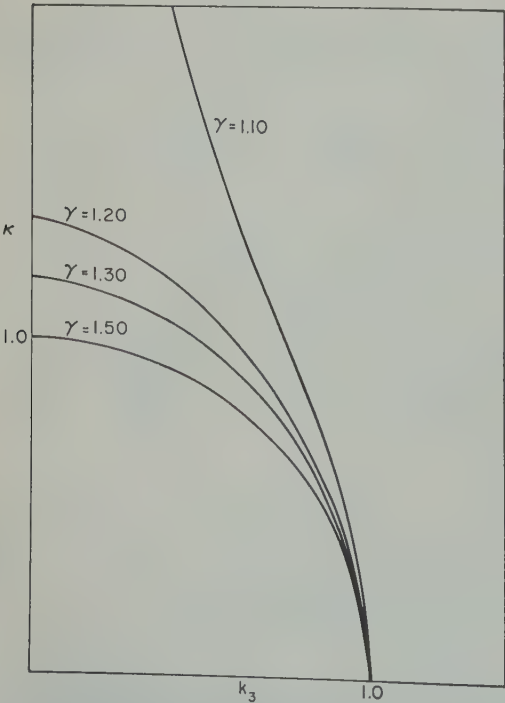


Fig. 11. Slowness surfaces in the lower cross-hatched region of Figure 5.

$$Q = \frac{1}{\rho_0} \frac{[ik_3 + (g/C_0^2)]}{N^2 - \omega^2} p$$

where p and Q are complex numbers with relative phase given by

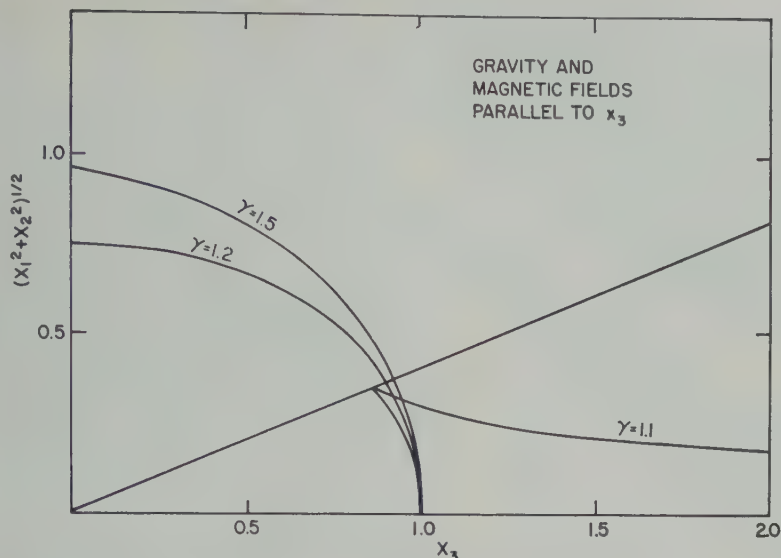


Fig. 12. Constant phase surfaces corresponding to the slowness surfaces of Figure 11.

$$\tan^{-1} \frac{k_3 C_0^2}{g}$$

variation in entropy leads the variation in pressure for $\omega < N$. The variation in pressure lags the entropy provided $\omega > N$.

The J.W.K.B. approximation gives a qualitative picture of wave propagation where the density of the medium varies owing to the presence of the gravitational field. The parameters C_A , C_0 , and N_1 will be functions of position so that the dispersion relation is also a function of the spatial coordinate. The wave form calculated is an approximation for a time short compared with the scale length divided by the group velocity. The considerations indicate the importance of the parameters N and N_1 to the propagation of low frequency hydromagnetic waves. The propagation is dispersive at these low frequencies. At frequencies less than ω_2 a new mode appears, incorporating some of the characteristics of the T mode and the P mode, and having some resemblance to the classical gravity wave [Eckart, 1960].

3. PERPENDICULAR MAGNETIC AND GRAVITATIONAL FIELDS

The effect of a gravity field is to couple the T and P modes at low frequencies. If the mag-

netic and gravitational fields are parallel, then one mode is still separable, provided new variables, vorticity, and two-dimensional divergence are chosen; no new choice of variables leads to decoupling in the general case. Once the gravitational and magnetic fields are at an angle, the V mode is no longer separable. In this case, the vorticity about the line of force involves a component of velocity along the direction of the gravity. The component of velocity along the gravitational field implies pressure and entropy variations. The transverse mode also involves a component of velocity along the gravitational field. A further complication is that the axial symmetry is lost, and the slowness surfaces are triaxial.

The effect of a gravity field at an arbitrary angle to the line of force is studied by taking the lines of force parallel to the x_1 and $g \partial z / \partial x_1$ parallel to x_3 . The distribution of density is assumed to be the result of gravitational forces only. The relevant equations are then

$$\begin{aligned} \frac{\partial b_1}{\partial t} &= -\frac{\partial v_2}{\partial x_2} - \frac{\partial v_3}{\partial x_3} \\ \frac{\partial b_2}{\partial t} &= \frac{\partial v_2}{\partial x_1} \\ \frac{\partial b_3}{\partial t} &= \frac{\partial v_3}{\partial x_1} \end{aligned} \quad (25)$$

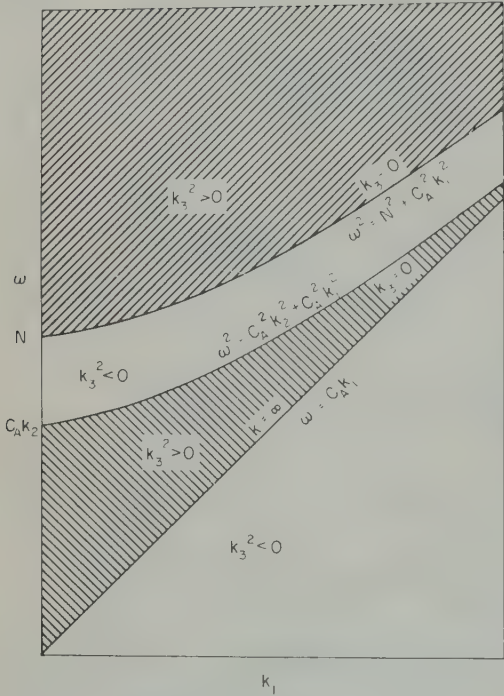


Fig. 13. Diagnostic diagram for $\beta = 0$; gravity and magnetic field at right angles. $\omega - k_1$ projection at a fixed value k_2 with $C_A k_2 < N$.

$$\frac{\partial v_1}{\partial t} = -\frac{1}{\rho_0} \frac{\partial p}{\partial x_1}$$
$$\frac{\partial v_2}{\partial t} = -\frac{1}{\rho_0} \frac{\partial p}{\partial x_2} - C_A^2 \left[\frac{\partial b_1}{\partial x_2} - \frac{\partial b_2}{\partial x_1} \right]$$
$$\frac{\partial v_3}{\partial t} = -\frac{1}{\rho_0} \frac{\partial p}{\partial x_3} - C_A^2 \left[\frac{\partial b_1}{\partial x_3} - \frac{\partial b_3}{\partial x_1} \right]$$
$$+ N^2 Q - \frac{1}{\rho_0} \frac{g}{C_0^2} p$$
$$\frac{\partial p}{\partial t} - g \rho_0 V_3 + \rho_0 C_0^2 \left[\frac{\partial v_1}{\partial x_2} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3} \right] = 0$$
$$\frac{\partial Q}{\partial t} + v_3 = 0$$

Before considering the dispersion relation for the complete set of equations, we examine equations 25–28 under the approximations $\beta \ll 1$; $\alpha \approx \gamma$. The dispersion relations give

$$k_3^2 = \frac{[(\omega^2 - N^2) - C_A^2 k_1^2][\omega^2 - C_A^2(k_1^2 + k_2^2)]}{C_A^2(\omega^2 - C_A^2 k_1^2)}$$

(29)

$$k_2^2 = \frac{[\omega^2 - C_A^2 k_1^2][\omega^2 - N^2 - C_A^2(k_1^2 + k_2^2)]}{C_A^2(\omega^2 - N^2 - C_A^2 k_1^2)}$$

(30)

where the loss of symmetry is evident. We note that for $\omega \gg N$ the propagation is a spherically spreading Alfvén wave.

Since the propagation is asymmetric, it is useful to consider the diagnostic diagram in the $\omega - k_1$ projections. Figure 13 illustrates the $\omega - k_1$ (eq. 29) projection of the diagnostic diagram. The shaded regions correspond to those regions in which k_3^2 is greater than zero for a fixed value of k_2 . The lower region is bounded by the line $\omega = C_A k_1$. The upper boundary is given by

$$\omega^2 = C_A^2(k_2^2 + k_1^2)$$

(31)

provided that

$$k_2 < N/C_A$$

At large values of the wave number k_1 , the upper boundary of the lower region of allowed propagation asymptotically approaches the line $\omega = C_A k_1$. This line corresponds to a one-dimensional Alfvén wave propagation. In the lower

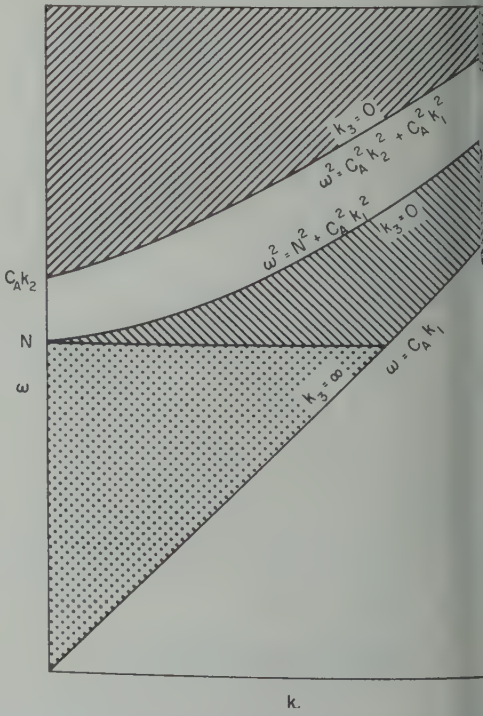


Fig. 14. Diagnostic diagram for a fixed value k_2 with $C_A k_2 > N$.

region, k_3 varies from zero to infinity, in this region the perturbed V mode spreads out the line of force because of the action of gravity. Furthermore, the curvature of the propagation lines indicates that the propagation is dispersive.

The upper region is bounded by the curve

$$\omega^2 = N^2 + C_A^2 k_1^2 \quad (32)$$

which asymptotically approaches $\omega = C_A k_1$ for large values of k_1 . The propagation in this region corresponds to a perturbed T mode. If N/C_A , the diagnostic diagram takes the form given in Figure 14. The lower region of allowed propagation is divided into two at $\omega = N$. The upper boundary of the lower region is given by equation 32 and the lower boundary is given by equation 31. As for N/C_A , the bounding curves asymptotically approach $\omega = C_A k_1$.

The diagnostic diagram corresponding to equation 30 is shown in Figure 15. Taking k_3 as a parameter, the allowed regions in the $k_1 - k_2$ space are projected; k_2 becomes large along the line given by equation 32 and vanishes along the line $\omega = C_A k_1$ and along

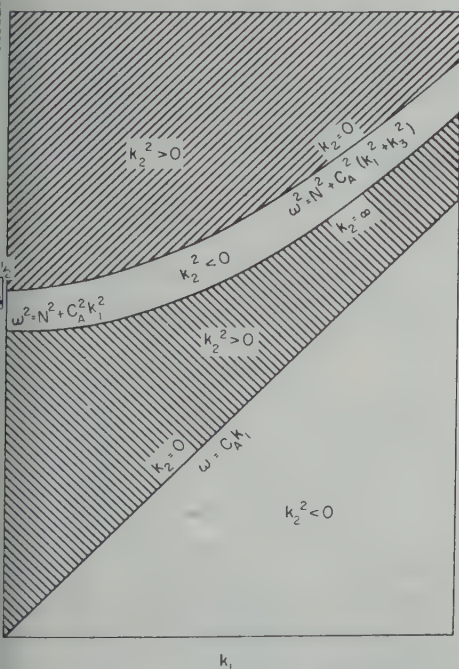


Fig. 15. Diagnostic diagram for a perpendicular magnetic and gravity field at a fixed value of k_3 .

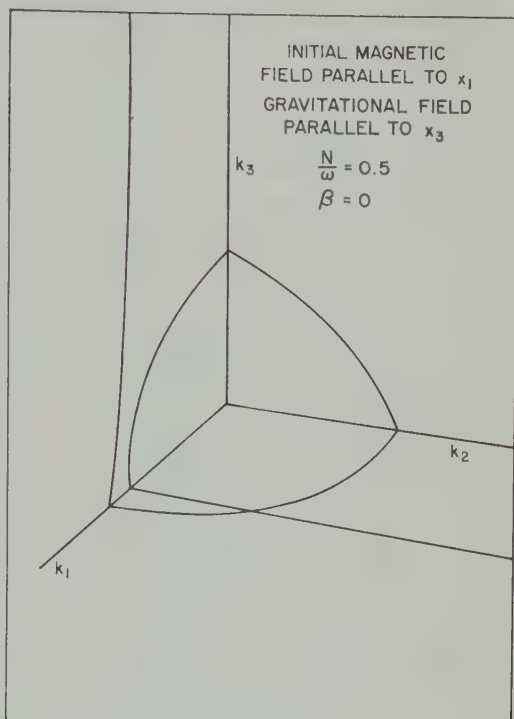


Fig. 16. Slowness surfaces for perpendicular gravity and magnetic field at $\beta = 0$. A complete surface is obtained by reflection across the coordinate planes.

the curve fixed by equation 32. The propagation in the upper region will be nearly spherical while the propagation in the lower region corresponds to a one-dimensional wave spreading about the line of force. As k_3 becomes small, the lower region of allowed propagation shrinks to a double valued boundary line.

The representation of the slowness surfaces corresponding to the diagnostic diagrams in Figures 13-15 is difficult because of the essential three-dimensional character of the propagation. Furthermore, the representation of the constant phase surfaces with several cuspidal edges is not easily managed. The equation of the slowness surface is

$$k_1^4 + [k_2^2 + k_3^2 - l^2(2 - \alpha^2)]k_1^2 + l^2[l^2(1 - \alpha^2) - (k_2^2 + k_3^2) + \alpha^2 k_2^2] = 0 \quad (33)$$

where α is given by $\alpha = N/\omega$; $l = \omega/C_A$. Figure 16 gives a schematic representation of the slowness surface for $N/\omega = 0.5$.

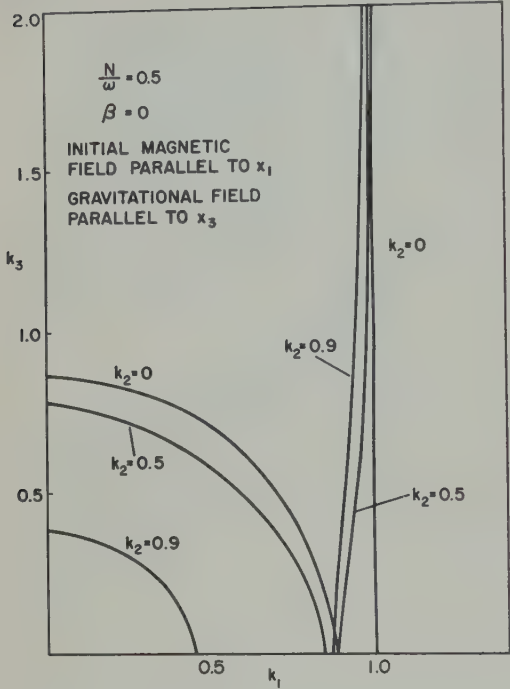


Fig. 17. The slowness surface for $N/\omega = 0.5$. The surfaces are projected onto $k_1 - k_3$ plane. The near plane corresponds to perturbed V mode propagation, the sphere to T mode propagation.

The two sheets of the propagation surface are shown in Figures 17 and 18. Figure 17 is a projection of propagation surface equation 33 onto the $k_1 - k_3$ plane. The complete surfaces are obtained by reflection across the coordinate planes. The lower region of the diagnostic diagram Figure 13 collapses into a line for $k_2 = 0$, and the propagation of the V mode is one-dimensional at all frequencies. The upper region corresponds to a spherically spreading mode, a perturbed T mode. The $k_2 = 0.5$ cut shows that the V mode has now spread about the k_1 axis; there is a parabolic point in this plane and a corresponding cuspidal edge in the phase surface. The T mode is a triaxial ellipsoid as is evidenced from the $k_2 = 0.5, 0.9$ cuts through the phase surface. As k_2 becomes larger, the V mode slowness surface shows a greater curvature.

Figure 18 is the projection of the slowness surface onto the $k_1 - k_2$ plane. For large values of k_3 the propagation becomes nearly one dimensional along the k_1 axis and there is no T mode. We note that at $k_2 = 0$ the two sheets of the

slowness surface intersect (see Figs. 15 and 16). At lower frequencies ($\omega < N$) there is only a single propagating mode. The projection of the slowness surface on the $k_1 - k_3$ plane is shown in Figure 19. The propagation is one dimensional at Alfvén wave velocity along the line of force for $k_2 = 0$. At higher values of the wave number k_3 , the propagation spreads about the k_1 axis, so that for large values of k_2 the propagation is along the x_3 axis. In the presence of a gravitational field, the propagation at very low frequencies takes the form of waves spreading about the direction perpendicular to the lines of force. The $k_2 - k_3$ plane cuts the slowness surface along hyperbola whereas the $k_1 - k_3$ projection is a modified circle.

The dispersion relation for finite β is

$$\begin{aligned} &\beta^2 k_1^6 + \{\beta^2[k_2^2 + k_3^2 - (2 - \alpha^2)l^2] - l^2\} \\ &+ [\beta^2\{l^4 - l^2[\alpha^2(l^2 - k_2^2) + 2(k_3^2 + k_2^2)] \\ &+ l^4(2 - \gamma^2) - l^2(k_3^2 + k_2^2)]k_1^2 \\ &+ \beta^2 l^4(k_3^2 + k_2^2 - \alpha^2 k_2^2) \\ &+ l^4[k_3^2 + k_2^2(1 - \gamma^2) + l^2(\gamma^2 - 1)] = 0 \end{aligned}$$

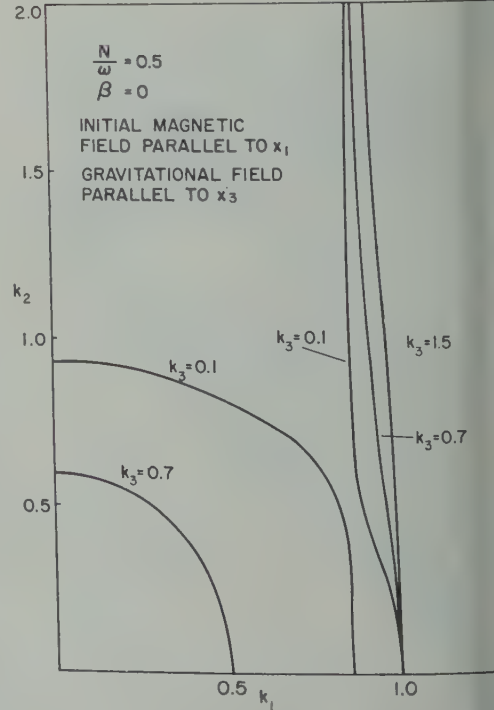


Fig. 18. The slowness surfaces of Figure 17 shown on a $k_1 - k_2$ projection.

corresponding slowness surfaces show features similar to both those of Figures 6 and 8 of Figures 17, 18, and 19. The slowness surfaces are composed of three sheets: two corresponding to nearly one-dimensional propagation and a third, a triaxial ellipsoid, corresponding to the propagation of the gravitationally perturbed T mode. The projections on the $k_1 - k_2$ plane of the sheets for $\beta = 0.5$ and $\gamma = 0.5$ are shown in Figure 20. The projection on the $k_1 - k_2$ plane shows the two nearly one-dimensional P and V mode propagations. The T mode propagation for $k_2 = 0$ is strictly one-dimensional along the direction of the magnetic field. The slowness surfaces show curves of nodal points both in the V and P modes, which transform into cuspidal edges in the constant phase surfaces.

The transition between the perturbed P - V - T modes to the gravitational modes for finite β is shown in Figures 21-24. For $\gamma = 0.9$ the triaxial ellipsoid representing the perturbed T mode has a cusp to a surface about the origin; the cor-

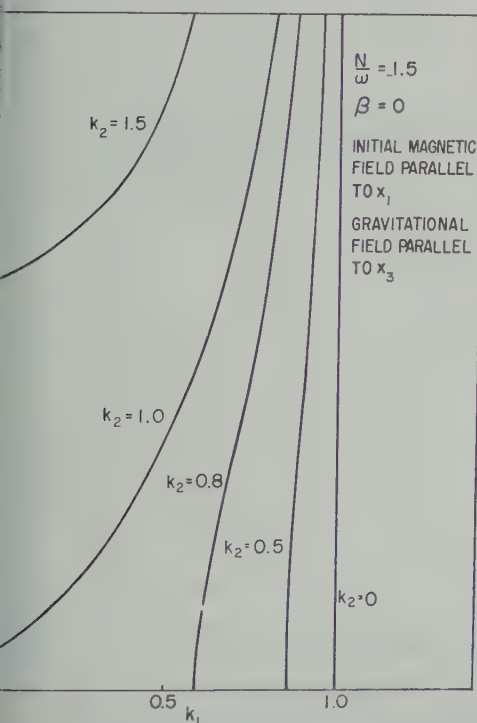


Fig. 19. Slowness surface for propagation at $N/\omega = 1.5$. The three-dimensional slowness surface is an asymmetric hyperbola.

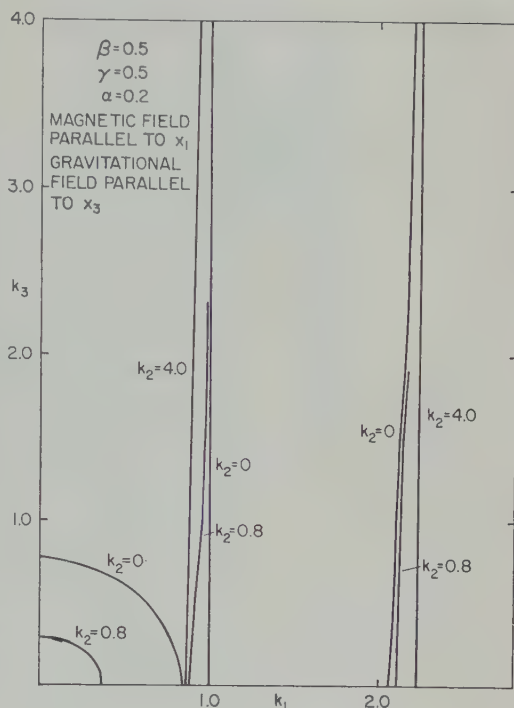


Fig. 20. The slowness surface for a perpendicular gravitational and magnetic field at $\beta = 0.5$. Slowness surfaces consist of three sheets corresponding to the perturbed P , V , and T modes.

responding phase velocity is higher than in the absence of gravity. The phase velocity of the perturbed V mode in the k_1 direction increases with increasing γ while the phase velocity of the perturbed P mode decreases.

At the transition frequency, $\gamma = 1.0$, there is no propagation in the T mode and only the one-dimensional highly perturbed V and P modes remain. The slowness surface for the P mode remains very nearly a plane at all values of the wave number, k_2 , but the V mode propagation undergoes spreading at intermediate values of the wave number, k_2 . As the frequency decreases, the V mode is propagated along the direction perpendicular to the lines of force, and the much perturbed acoustic mode (P) remains one-dimensional along the direction of a magnetic field.

Gravity renders the propagation dispersive. In the presence of perpendicular magnetic and gravitational fields the constant phase surfaces corresponding to the gravitationally perturbed P and V modes show cuspidal edges correspond-

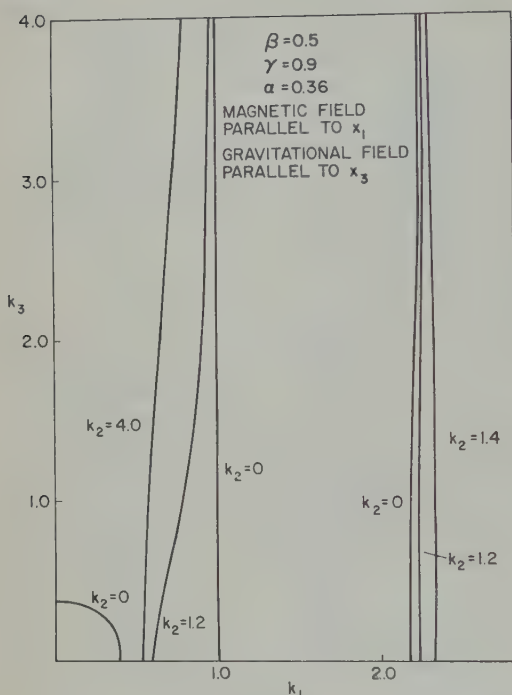


Fig. 21. Slowness surfaces for perpendicular gravity and magnetic fields at frequency just below the critical frequency N_1 .

ing to the curves of parabolic points on the slowness surfaces. The disturbance tends to be concentrated at the cuspidal edges and the conical point, so that a planar projection of the disturbance would show a disturbance in the form of an ellipse, with a central point. The dispersion is such that for frequencies less than N_1 the lower frequency components travel at a greater phase velocity than the higher frequency components. The effect of gravity in the case of both parallel and perpendicular gravity and magnetic fields is to couple the modes in a complicated way at frequencies less than the critical gravitational frequency N_1 .

4. CONTINUUM APPROXIMATIONS

The exosphere is a low density, partly ionized gas. A description of waves in the exosphere by the equations of magnetohydrodynamics implies that the gas can be treated as a continuum. In dealing with a rarified neutral gas the continuum equations would apply, provided that the length scale of the phenomena were large compared with the mean free path. The wide range of

physical phenomena involved in an ionized gas does not permit a simple working rule defining the limit of applicability of the continuum equations. The problem is particularly vexing in the exosphere, since conventional criteria indicate that a wide range of processes does not satisfy the conditions of either a continuum or a simple kinetic theory. An exploration of the problem would require a development of the predictions of the continuum theory and the comparison of these predictions with experiment. As an example, the continuum theory makes certain suggestions regarding the effect of a satellite on the exosphere. Experimental tests of these predictions would allow an estimate of the applicability of the continuum theory to other phenomena having comparable length scales.

The sound wave velocity, $C_0 = \sqrt{(\gamma p/\rho)}$, is about 1 to 3 km/sec in the exosphere. The Alfvén wave velocity C_A is much larger, ranging from several hundred to several thousand kilometers per second. The ratio of these two velocities, C_0/C_A , is a small number, on the order of 10^{-3} – 10^{-4} . The ratio of the charged particle pressure to

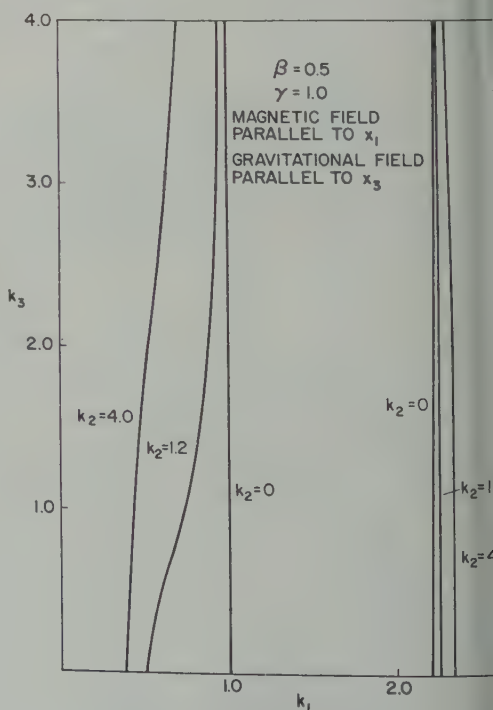


Fig. 22. The slowness surfaces for a perpendicular gravity and magnetic field at the critical frequency N_1 .

etic pressure, $H^2/8\pi$, is of order β^2 . Thus, the entire exosphere, the magnetic pressure dominates the particle pressure. It should be noted that the usual interpretation of the Alfvén velocity as a speed of propagation of transverse waves along magnetic lines of force in an incompressible fluid requires that the Alfvén wave velocity be much less than the sound wave velocity. When the opposite holds true, then the Alfvén waves must be reinterpreted as one-dimensional propagations of vorticity and the transverse propagation of the transverse components of velocity.

The P mode involves the propagation of perturbations in charged particle pressure. It would seem that the characteristic length limiting the continuum approximation is the length over which an appreciable proportion of particle momentum is exchanged. The V and T modes involve the propagation of the transverse components of the velocity. Because of the magnetic field, the charged particles spiral about the lines of force. The motion of the charged particles in

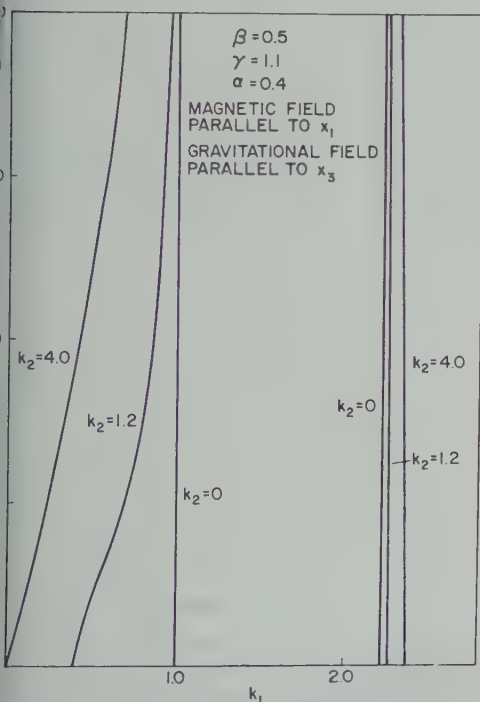


Fig. 23. The slowness surface for perpendicular magnetic and gravitational fields at $\gamma = 1.1$. The perpendicular P mode remains almost one-dimensional. The V mode becomes highly distorted. The T mode propagation is highly dispersive.

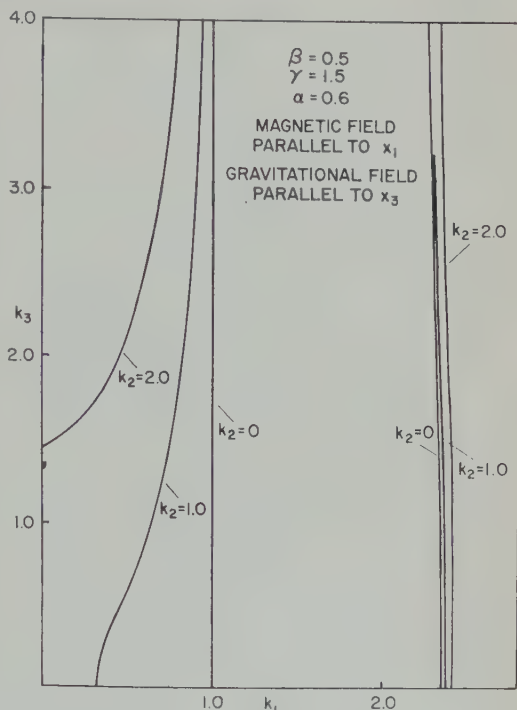


Fig. 24. Propagation of low frequency for perpendicular magnetic and gravitational fields.

directions at right angles to the lines of force is greatly restricted, while the motion along the lines of force is largely independent of the magnetic field. We would expect, then, that the characteristic length signifying applicability of the continuum approximations would involve the radius of gyration. The fact that different length scales are involved in the different modes is intimately connected with the anisotropic nature of the wave propagation. Thus, there may be conditions for which V and T mode propagation may take place, yet there would be no P mode propagation.

In a partly ionized gas there are several mean free paths corresponding to neutral-neutral collisions, neutral-charged particle collisions, etc. Inelastic processes, such as ionization, recombination, and charge exchange, introduce further complications. The neutral-neutral mean free path is the distance a neutral particle must travel before it exchanges its momentum by a direct collision with another neutral molecule. The difference between the neutral-neutral mean free path and the neutral-charged particle mean free path is small. Over the entire exosphere the

neutral gas mean free path is in excess of several kilometers and is large compared with other lengths (except the scale height) that characterize the motion of charged particles. It is because of this large difference in the representative length scale of neutral and charged particle gas that it is usually assumed that the two gases act independently. However, for large disturbances in the lower exosphere this no longer holds true, and the influence of the neutral molecules on the wave propagation must be considered.

As a result of the Coulomb interactions, a charged particle exchanges its momentum, not by single collisions, but in small bits, through distant interaction. Because of the small but frequent exchanges of momentum, a particle may transfer an appreciable portion of its momentum in distances short compared with the distance a particle must travel to exchange all of its momentum. In the calculation of the charged particle mean free path [Ferraro, 1933; Spitzer, 1956], the angular deflection in the path of the particle traveling at a velocity V relative to a given particle is computed as a function of the distance between the two. A particle traveling a distance r will, on the average, come to within a distance $s = 1/n_i \pi r^{1/2}$ where n_i is the charged particle density. The loss of directed momentum is obtained by summing up over encounters at different distances of separation. The lower limit of the integration is taken as zero corresponding to direct but highly improbable collisions, whereas the upper limit is taken as the Debye length $(kT/4\pi n_i e^2)^{1/2}$. The interpretation of the Debye length is the distance over which the removal of all charged particles of one sign would result in an electric field potential comparable to kT . The Debye length ranges from a few tenths to a few tens of centimeters in the exosphere. The distance an electron must travel in order to exchange all of its momentum is then given by

$$l_m = \frac{1}{4\pi n_i} \frac{(3kT)^2}{e^2} [\ln(1 + a^2)]^{-1}$$

where

$$a = \frac{3}{2} \frac{ze^3 (k^3 T^3)^{1/2}}{\pi n}$$

Alternatively, the distance a particle must travel to exchange a fraction x of its initial momentum $x^2 = V^2/\bar{V}_e^2$ is

$$\frac{1.29 \times 10^5 T^2}{\ln(a + 1)n_i} x^2$$

where V is the component of velocity perpendicular to the initial motion of the particle and \bar{V} is the thermal velocity. The quantity $\ln(a + 1)$ is about 10–20 in the exosphere [Spitzer, 1956]. Thus, the elementary theory of the particle interaction states that the distance a particle must travel to exchange a fraction x of its momentum depends on the square of the proportional change in momentum.

The interpretation of the V mode as the propagation of vorticity along the line of force implies that the condition for the continuous approximation to hold is that the disturbances have a length scale large compared with the rms radius of the spiral of the ions about the line of force. The Larmor radius is $(kTm_i)^{1/2}/eB$, where m_i is the mass of the charged particle. V can also interpret the Larmor radius as the length over which the particle pressure, nkT , is comparable to magnetic Lorentz force $Bne(kT/m)^{1/2}$. In a strong field, a dilute gas has a Larmor radius much smaller than the mean free path. Collisionless hydromagnetic waves are then propagated through the intervention of the external field.

We consider the applicability of the continuous equations for two models of the exosphere. In both of these models the magnetic field is taken to be that of the dipole with an equatorial surface field of 0.315 gauss. Two distributions of charged-particle density are considered. The high density model is that proposed by Dessler, Francis, and Parker [1960] and based on the ion-density values of Johnson [1960]. Johnson has calculated the distribution of ions in gravitational field above the F_2 layer assuming diffusive equilibrium. The model gives maximum ion number density of 1.1×10^6 particles at an altitude of 4.5×10^3 km and a density of 4 particles at an altitude of 2.5×10^4 km. The molecular weight of the lower exosphere is decreasing rapidly between the region of 1000 to 2500 km to 1.0. The assumed magnetic field and density distribution combine to yield the variation of Alfvén wave velocity shown as the high density model in Figure 25.

An alternative model of the exosphere has been constructed to indicate the range of possibilities. Hanson and McKibbin [1961] have

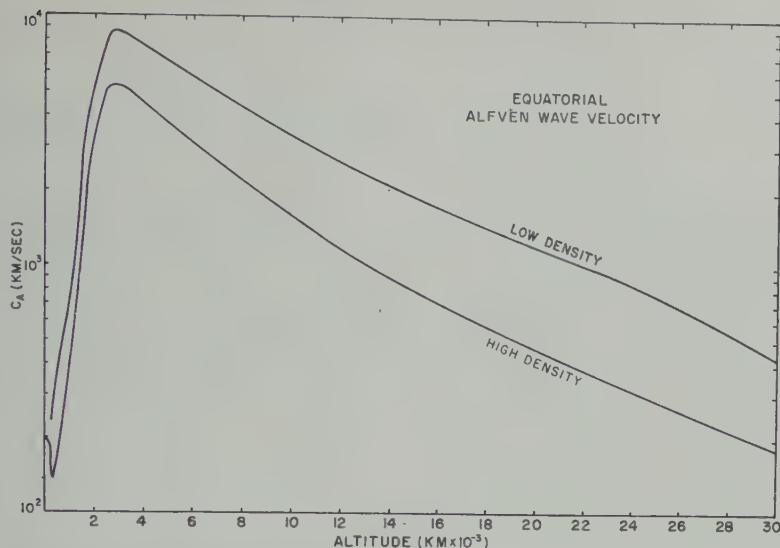


Fig. 25. The variation of equatorial Alfvén wave velocity for two models of the exosphere.

ed out an ion-trap measurement of the ion concentration profile in the region of 240 to 800 km in an early evening firing. Above 800 km the concentration profile approximates hydrostatic distribution of ions of mass 16 at temperature of 1240°K. *Smith and Helliwell* [1961] have obtained an electron density of 10¹⁰ electrons/cc at 5 earth radii using nose probes. This value is in good agreement with determination by *Blackwell and Ingham* [1961] of outer coronal densities. However,

Johnson and Fish [1960] and *Johnson* [1961] suggest that the whistler medium is of telluric origin. In this event the agreement in density of the solar corona and the whistler medium is fortuitous. The whistler values suggest somewhat lower density than those used by Dessler, Francis, and Parker. The low density model is based at large distances on the whistler data and in the lower exosphere on the rapid determinations of Hanson and McKibbin. The hydrostatic calculation at 1300°K has been carried out with assump-

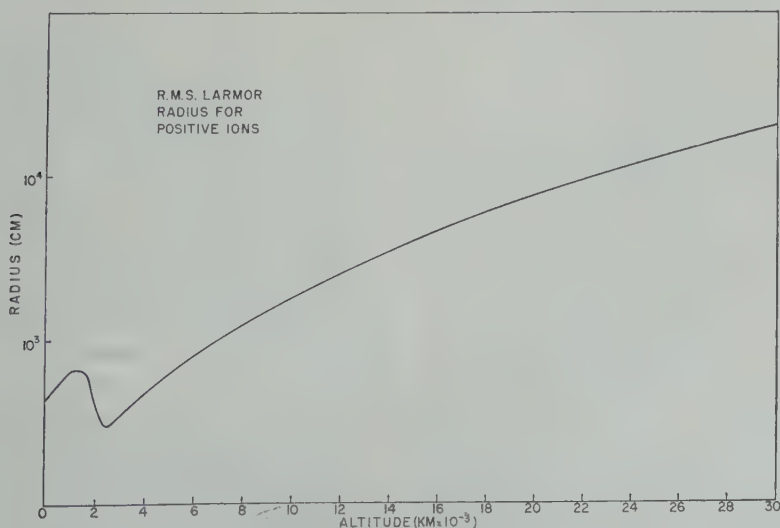


Fig. 26. The variation of the rms Larmor radius for positive ions in the exosphere at 1300°K.

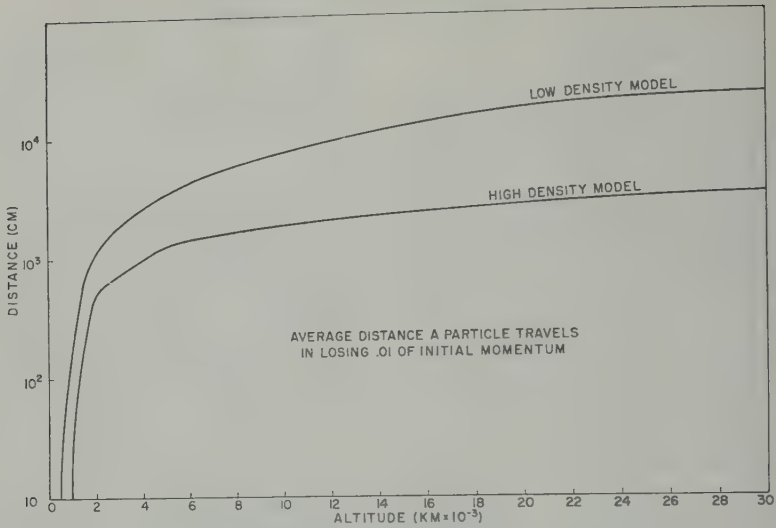


Fig. 27. The average distance a particle must travel in exchanging 1 per cent of its initial momentum for two models of the exosphere.

tions regarding the variation of ion composition similar to those used by Johnson. The magnetic field and density combine to give the low density model variations of the Alfvén wave velocity shown in Figure 25. The two models agree closely though they may represent widely different conditions. The major uncertainties in the velocity distribution is in the critical region of from 1000 to 3000 km altitude.

The rms Larmor radius depends only on the mean molecular weight, temperature, and magnetic field. For an isothermal exosphere of temperature 1300°K the two models lead to similar rms Larmor radius and this is shown in Figure 26. At an altitude of 30,000 km the rms Larmor radius is about 100 meters.

Figure 27 shows the approximate distance an electron travels in exchanging 1 per cent of its initial momentum. The mean free path is then 10^4 times the listed values. We note that in the outer exosphere, using the high density model, a charged particle travels about 1 km before exchanging 10 per cent of its momentum. In the lower exosphere, a particle exchanges 10 per cent of its momentum in traveling a distance on the order of meters.

The interpretation of Figures 26 and 27 is that for disturbances having a length scale large compared with the distances plotted, the continuum equations yield at least a qualitative description. At distances small compared to those

given, the problem should be treated either from the point of view of a Boltzmann theory or from a simple kinetic theory. It should be noted that the T and V modes, but not the P mode might be excited by disturbances having an intermediate length scale.

The applicability of the macroscopic equations has been dealt with in terms of length rather than in terms of frequency collision time. The characteristic length can be converted into characteristic frequency provided that the disturbance has a characteristic macroscopic velocity.

In summary, the propagation of the T and V modes can be treated by the continuum equations at least qualitatively, provided that the disturbance has a length scale long compared with some hundreds of meters in the outer exosphere and some tens of centimeters in the lower exosphere. The P mode propagation will be qualitatively described by the continuum approximation, provided that the length scale is large compared with tens of meters in the lower exosphere and tens of kilometers in the upper exosphere.

5. DESCRIPTION OF SOURCES

The gravitationally induced inhomogeneity of the charged particle density field does not enter into the description of a source, provided the length scale associated with the source

compared with the length scale of the inhomogeneities. Additional problems arise in a medium in that the inhomogeneities may cause mode conversion. In the present section we will consider the description of sources in a homogeneous medium.

The sources of the P and T modes differ in an important way from a source generating V waves. The source for the V mode must generate vorticity. A pressure fluctuation is propagated in both P and T modes. A pressure fluctuation may arise from either the local injection of charged particles or heat. An example in the exosphere is when a pressure source would be the influx of ionizing radiation into a neutral gas, or the motion of a satellite or missile through the charged particle gas.

The region of allowed propagation of the V wave in the diagnostic diagram is confined to a narrow band. This implies that the waves generated by a source are characterized by a limited range of wave numbers. The V wave will have a narrow band spectrum. The region of the T mode permits a broad spectrum. The P wave will be intermediate, with at small values of β the allowed region approaching a line. The character of the spectrum of magnetic disturbance can thus be used for qualitative identification of the V and T modes. A pressure source can be described in terms of its characteristic length L . A source with a characteristic velocity V has a characteristic frequency V/L . A qualitative picture of the distinction of T and P modes generated by a source having a velocity V and a length L can be obtained from a consideration of the diagnostic diagram in Figure 1.

At a given frequency, a source with a small length scale (high characteristic wave number) will tend to have the energy concentrated in the region of P mode propagation and, as a result, the source generates primarily P waves. At the same frequency a source with a large length scale will generate primarily T waves. Qualitatively, the characteristic length of the source should be compared with the characteristic wave length of the mode at a given frequency. This wave length is of the order $(C_0/V) C_0$ and $(L/V) C_A$ for the P and T modes respectively. Thus, for a satellite having a length of a meter and traveling at 8 km/sec, the characteristic wave length of the P mode is about 100 m, while the characteristic wave length of

the T mode is 300 meters. The satellite under these conditions would generate primarily P modes, provided that the continuum conditions are met.

The proportion of V waves that are generated by a given source depends on the strength of the vorticity source. Vorticity is generated by the breakdown of surfaces of discontinuity in the fluid flow. The surface of discontinuity can arise between fluids traveling at different velocities or when the fluid passes over a solid. The characteristic aspect of vorticity generation is that one part of the interior of the fluid is subjected to pressure for a time while the adjacent parts are not. Surfaces of discontinuity may arise at the boundary of the pressure pulse generated by incoming ionizing radiation or by the breakdown of the finite amplitude front associated with the plasma reaching the earth's field. For satellites and missiles, vorticity is generated by the exhaust in a region of partly ionized gas. The dimensions of a typical rocket suggest that a V wave is generated at a peak frequency of about 10^8 cps, provided the burnout is in or above the ionosphere. The resulting signal would not undergo geometric attenuation, but at high frequencies conductivity losses are large. An experimental verification is possible by observation at the conjugate to the launch site.

Vorticity may also generate in the wake of a satellite. The character of the wake is determined by the Reynolds number.

$$R = \frac{VL}{\nu} \cong \frac{VL}{C_0 L_M}$$

where ν is the kinematic viscosity and L_M is the mean free path. If the Reynolds number is much greater than 1, then the wake is turbulent [Townsend, 1956]. In the lower exosphere the Reynolds number is much less than 1 and the wake is laminar, with vorticity diffusing to form a broad disturbed region.

The above comments regarding the effectiveness of a satellite as a source of magnetohydrodynamic waves depends on whether the medium can be treated as a continuum for disturbances having a length scale of the dimensions of a satellite. At 300 km, the rms Larmor radius for the positive ions is of the order of a meter, and a particle must travel about 3 meters in exchanging 10 per cent of its momentum. Under these conditions we will expect that a satellite having

dimensions on the order of meters generates macroscopic V and T waves, but that a continuum treatment of the P mode is only approximate.

A quantitative theory of the relative proportions of P , V , and T waves generated by a given source depends on the detailed treatment of finite amplitude phenomena. An estimate of the relative proportions of P and T modes for a weak source is possible [Lighthill, 1960]. We consider a pressure source P_s ,

$$\frac{\partial p}{\partial t} + \rho_0 C_0^2 \left(\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3} \right) = P_s(x_i, t)$$

P_s has the units of dynes/cm²/sec. The pressure source is related to a mass source and to the net accession of heat by

$$P_s = \frac{\rho_0(\gamma_0 - 1)}{\alpha_0 T_0} q + C_0^2 S \quad (34)$$

where q has a unit of ergs/g/sec and S has the units of g/cm³/sec. The pressure source P_s will not generate vorticity.

Eliminating v_i and P from the field equations (14a-17a) we obtain

$$\begin{aligned} \frac{\partial^2 R}{\partial t^2} - C_0^2 \left[\frac{\partial^2 R}{\partial x_1^2} + \frac{\partial^2 R}{\partial x_2^2} + \frac{\partial}{\partial x_3} \right. \\ \left. \cdot \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right) \right] V_3 - C_A^2 \bar{V}^2 R \\ = -\frac{1}{\rho_0} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right) P_s \end{aligned} \quad (35)$$

$$\frac{\partial^2 v_3}{\partial t^2} - C_0^2 \left[\frac{\partial R}{\partial x_3} + \frac{\partial^2 v_3}{\partial x_3^2} \right] = -\frac{1}{\rho_0} \frac{\partial P_s}{\partial x_3} \quad (36)$$

where R is the two-dimensional divergence

$$R = \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2}$$

Equations 35 and 36 combine to yield

$$\begin{aligned} \frac{\partial^2}{\partial t^2} \left[\frac{\partial^2}{\partial t^2} - (C_0^2 + C_A^2) \bar{V}^2 \right] R \\ + C_0^2 C_A^2 \frac{\partial^2}{\partial x_3^2} \bar{V}^2 R \\ = -\frac{1}{\rho_0} \frac{\partial^2}{\partial t^2} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right) P_s \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial t^2} \left[\frac{\partial^2}{\partial t^2} - (C_0^2 + C_A^2) \bar{V}^2 \right] \frac{\partial v_3}{\partial x_3} \\ + C_0^2 C_A^2 \bar{V}^2 \frac{\partial^3}{\partial x_3^3} v_3 \\ = -\frac{1}{\rho_0} \left[\frac{\partial^2}{\partial t^2} - C_A^2 \bar{V}^2 \right] \frac{\partial^2}{\partial x_3^2} P_s \end{aligned}$$

The space-time Fourier transform of the source term for the two-dimensional divergence is the

$$-\frac{1}{\rho_0} \omega^2 (k^2 - k_3^2) \text{ F.T. } (P_s)$$

and the corresponding Fourier transform of the source for $\partial v_3 / \partial x_3$ is

$$-\frac{1}{\rho_0} (\omega^2 - C_A^2 k^2) k_3^2 \text{ F.T. } (P_s)$$

The source terms can also be written in terms of the net accession of mass and heat using equation 34.

A point in space and time has a constant Fourier transform of P_s , all wave numbers and all frequencies being equally represented in a source. A source that is a point in time but isotropic in space with a length scale L may be represented by

$$\text{F.T. } (P_s) = \frac{1}{8\pi^3} e^{-(1/4k^2 L^2)}$$

corresponding to the gaussian pressure field distribution about the origin. A source with characteristic frequency ω would include a function at that frequency.

The source terms (37 and 38) make the greatest contribution to a wave at those values of wave number and frequency that lie on the slowness surface. The slowness surface for the P mode is approximately $\omega = C_A k$ and the source term for the P mode (eq. 38) vanishes. Along the slowness surface for the P mode

$$k_3^2 \approx \omega^2 C_0^{-2} [1 + \beta^2]$$

so that the greatest contributions from the source term of the P mode result when

$$\frac{C_A^2 k^2}{\omega^2} \gg 1$$

Since $k^2 e^{-k^2 L^2 / 4}$ has a maximum at $k = (2/L)^{1/2}$, the greatest contribution of the source term to the P mode is for

$$\frac{C_A^2}{\omega^2 L^2} \gg 1$$

all sources at low frequencies thus contribute most heavily to the P mode. The contribution of the source (eq. 37) to the T mode is at a value of k_s , is greatest for $k = (2L)^{-1}$ and high frequencies.

The representation of the source given above in a sense, conventional and applicable only if the discontinuity in pressure is not too great. If a large discontinuity, the source region will be turbulent. Turbulent interaction will lead to a complex source function. In addition to the direct generation of P and T modes by the action of pressure or entropy sources, vorticity sources generate the P mode, as can temperature inhomogeneities.

As an example of some of the difficulties involved in treating strong sources, we consider the Argus experiment [Christofilos, 1959; Newman, 1959]. In these experiments approximately 100 megajoules of energy were released at a height of about 450 km. The time behavior of the source is well represented by a δ function, but there is no unique length scale. We might, for instance, consider the length as determined by distance over which the initial pressure pulse must spread in order to attain the ambient pressure. The initial pressure of the charged particles is on the order of 10^{-7} dynes/cm², so that the indicated length scale would be of the order of 1000 km. This length is large compared with the variation in charged particle density, so that the source cannot be considered as an isotropic source. A source with a length scale of 1000 km could generate T waves with a peak frequency of a few tenths of a cycle per second. The observed first pulse appears to have been peaked around a cps [Newman, 1959]. In addition, the source should have generated low-frequency waves. No attempt was made to detect such waves. The turbulence associated with the outgoing shock should have generated vorticity leading to V wave propagation. There is some indication that these waves were excited since a sinusoidal wave form of about a cycle per second was observed at the conjugate point to the explosion.

Verthold, Harris, and Hope [1960] find that the amplitude of the low-frequency magnetic field generated by Argus was greatest along

the geomagnetic meridian. The signal strength falls off rapidly as the direction of propagation deviates from the geomagnetic meridian. Such an amplitude pattern is consistent with V mode propagation without geometrical attenuation along the line of force into the ionosphere. The propagation in the atmosphere was then a radially spreading electromagnetic wave, badly disturbed by the partly conducting lower and upper boundaries of the atmosphere.

6. SPECTRUM OF MAGNETIC DISTURBANCES

The spectrum of a hydromagnetic disturbance will be determined primarily by the source characteristics. The high-frequency cutoff is fixed by the Larmor radius at the source [Dessler and Parker, 1959]. The Larmor radius in the lower exosphere is of the order of meters and the corresponding cutoff frequency for the V and T modes is on the order of 100 cycles per second. Higher frequency waves will be propagated as modified electromagnetic waves. In the outer exosphere the Larmor radius is on the order of 10^4 centimeters and the corresponding cutoff frequency is on the order of a few cps. Thus, the high frequency cutoff of a magnetic disturbance can serve as an indication of source location. The ground level distinction between waves that have propagated as electromagnetic waves from source to receiver and those that have traveled as hydromagnetic waves to the base of the ionosphere and then as electromagnetic waves can be made on the basis of travel times. The source dimensions and time scale will determine the character of the initial onset. A source with a large-length scale will principally generate high-frequency T mode, some low-frequency P mode, and the proportion of V mode will depend on the sharpness of the discontinuity. The V mode may be identified by its characteristic narrow-band spectrum.

At low frequencies it is possible that hydromagnetic modes modified by gravity become important. The characteristic gravitational frequency N_1 as a function of altitude is shown in Figure 28. In the lower exosphere this characteristic frequency is of the order of 10 minutes, while in the upper exosphere the characteristic frequency is on the order of 15 hours. For frequencies large compared with this characteristic frequency, the propagation is in the unperturbed P , T , and V modes. There will be a gap in the

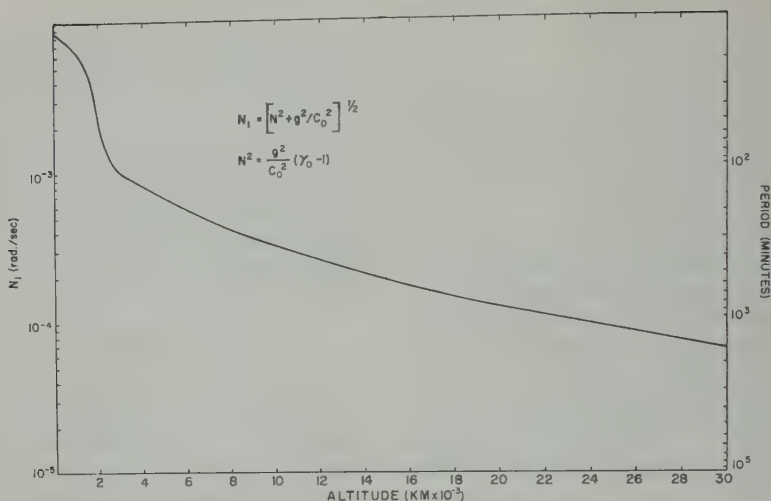


Fig. 28. Variation of the critical frequency N_1 as a function of altitude.

spectrum at those frequencies separating the gravitational modes from the transverse modes (see Figs. 5, 13, 14, and 15).

The presence of the boundaries in the system opens up the possibility of trapping of the energy in a given mode. At the lower boundary of the exosphere, there is a rapid variation in the charged-particle density. However, there is no evidence for an upper boundary; the charged particle distribution grades into the interplanetary or coronal density. It thus appears unlikely that the T mode can be trapped in a resonant cavity formed by the lower boundary of the exosphere and the disordered outer regions where the earth's magnetic field ends and the interplanetary field begins.

The P mode will be propagated one-dimensionally along the lines of the force since β is on the order of $10^{-3} - 10^{-8}$ over the entire exosphere. The propagation is similar to the propagation in a pipe with a bulge in the middle. Because of the density contrast in the F_2 regions, the P modes can be trapped along the line of force, and it is possible that a free oscillation or standing wave could be set up. The period of such a standing wave would be on the order of several hours to a day. These long period vibrations require a consideration of the rotational effects, particularly the Coriolis force and these have not been included in the above formulation. In addition, viscosity is of considerable importance since it varies as the mean free path, and the mean free path differs by a factor of

10^8 between the lower and upper exosphere. There is thus the possibility that a free oscillation in the P mode may be set up, though rotational and viscous effects will be of great importance.

The V mode propagates vorticity along the line of force. The rapid variation of charged particle density at both ends of the line of force suggests that energy could be partially trapped, leading to a free oscillation in the V mode. Since the charged-particle density varies continuously, the mode will be a leaky mode with energy passing through and radiating out as an electromagnetic wave at the end of the line of force. A rough estimate is obtained by taking the wave impedance as proportional to $\rho^{1/2}$. The density contrast in the ionosphere then suggests a factor of 10^4 for this mode of oscillation on the order of 10^4 neglecting all internal losses.

A direct calculation of the free period of the mode of oscillation is straightforward. The equation for the propagation of vorticity

$$\frac{\partial^2 \xi}{\partial t^2} - C_A^2(x) \frac{\partial^2 \xi}{\partial x^2} = 0$$

is a one-dimensional wave equation with a variable phase velocity. x is the distance along the line of force. The boundary conditions depend on the details of the physical process at the two ends of the lines of force. If the proportion of energy leaking out of the mode is small, then the boundary condition is that the vorticity vanishes. The altitude at which vorticity vanishes

fixed by the condition that the charged-particle collision frequency be high compared with the Larmor frequency. It should be noted that the vanishing of the vorticity is the vanishing of a current along the line of force. Under these conditions, the eigen frequencies of a uniform system are $f_n = nC_A/2L$, where L is the length of the line of force and n is the number of the overtone.

Calculations have been carried out to determine the free periods for V mode oscillations using a variable C_A employing the two models of the exosphere. The results are shown in Tables 1 and 2 where the periods are listed as a function of geomagnetic latitude. These calculations do not take into account the possible leaking of energy out the end of the line of force.

The sharp maximum in the Alfvén wave velocity at altitudes of about 3000 km (Fig. 25) allows the trapping of energy between this maximum and the F_2 layer. The characteristic periods are listed in Table 3. The periods are shorter than the periods of oscillation along the line of force.

A weak source located below the 3000-km maximum would primarily excite the shorter period oscillations. Disturbances in the upper exosphere associated with day-night effects may provide such a source. A sufficiently strong disturbance in the upper exosphere would be expected to excite both the short- and long-period oscillations.

A detailed comparison between the computed and observed periods is difficult because the data presented, for example, for Benioff [1960] and for Troitskaya [1961] is in the form of spectrogram records. The obvious complexity of the data precludes a zero-crossing analysis. The spectrogram of Tepley [1961] shows micropulsations

TABLE 2. Free Periods of Oscillation in V Mode for Low Density Model (in seconds)

Geomagnetic Latitude	Mode Number			
	$n = 1$	$n = 2$	$n = 3$	$n = 4$
30	12.1	10.3	4.85	4.39
35	9.98	9.01	3.99	3.69
40	8.77	8.04	4.11	3.24
45	8.63	7.34	6.00	3.09
50	13.9	6.95	6.59	4.98
55	29.2	10.9	7.10	6.28
60	74.4	26.4	16.6	12.1

with a frequency of 5 to 0.5 cycles per second together with higher frequencies. These frequencies are higher than one would expect for a free-mode oscillation at the geomagnetic latitude of Palo Alto.

Acknowledgments. I am indebted to A. J. Dessler, R. Holzer, N. F. Ness, and C. E. Palmer for critically reviewing the manuscript and suggesting a number of important changes. I wish to thank Clarence Wade for aid in numerically evaluating the numerous polynomials. William Cahill and Dwight Sumida carried out the calculations of the free periods of the V mode.

REFERENCES

- Bazer, J., and O. Fleischman, Propagation of weak hydromagnetic discontinuities, *Phys. Fluids*, **2**, 366-378, 1959.
 Behr, A., and H. Siedentopf, Untersuchungen über Zodiakalllicht und Gegenschein nach lichtelek-

TABLE 3. Periods of Oscillation for Energy Trapped Between F_2 and Maximum in C_A Curve (in seconds)

Geomagnetic Latitude	Mode Number			
	$n = 1$	$n = 2$	$n = 3$	$n = 4$
High Density Model				
35	10.8	4.75	3.09	2.30
40	9.38	4.11	2.67	1.98
45	8.42	3.67	2.38	1.77
50	7.68	3.35	2.17	1.61
55	7.14	3.11	2.01	1.50
60	6.73	2.93	1.89	1.41
Low Density Model				
40	7.53	3.10	1.98	1.46
45	6.66	2.75	1.75	1.29
50	6.05	2.49	1.59	1.17
55	5.65	2.31	1.47	1.08
60	5.26	2.16	1.38	1.02

TABLE 1. Free Periods of Oscillation in V Mode for High Density Model (in seconds)

Geomagnetic Latitude	Mode Number			
	$n = 1$	$n = 2$	$n = 3$	$n = 4$
30	14.2	12.8	6.17	5.65
35	12.0	11.1	5.47	4.85
40	11.0	9.98	7.01	4.35
45	14.5	9.16	8.69	5.39
50	29.1	11.0	8.46	8.12
55	67.3	24.3	15.3	11.2
60	184.6	64.8	40.2	29.4

- trischen Messungen auf dem Jungfrauojoch, *Z. Astrophys.*, **32**, 19-50, 1953.
- Benioff, H., Observations of geomagnetic fluctuations in the period range 0.3 to 120 seconds, *J. Geophys. Research*, **65**, 1413-1422, 1960.
- Berthold, W. K., A. K. Harris, and H. J. Hope, World wide effects of hydromagnetic waves due to Argus, *J. Geophys. Research*, **65**, 2233-2240, 1960.
- Blackwell, D. E., and M. F. Ingham, Observations of the Zodiacal light from a very high altitude station, II, Electron densities in interplanetary space., *Monthly Notices Roy. Astron. Soc.*, **122**, 129-141, 1961.
- Chapman, S. and V. C. A. Ferraro, A new theory of magnetic storms, *Terrest. Magnetism and Atmospheric Elec.*, **36**, 77-97, 171-186, 1931.
- Chapman, S., and V. C. A. Ferraro, A new theory of magnetic storm, II, The main phase, *Terrest. Magnetism and Atmospheric Elec.*, **38**, 79-86, 1933.
- Chapman, S., and V. C. A. Ferraro, The theory of the first phase of a geomagnetic storm, *Terrest. Magnetism and Atmospheric Elec.*, **45**, 245-268, 1940.
- Christofilos, N. C., The Argus experiment, *J. Geophys. Research*, **64**, 869-876, 1959.
- Carstou, J., Hydromagnetic waves in a compressible fluid conductor, *Proc. Nat. Acad. Sci.*, **46**, 131, 136, 1960.
- Dessler, A. J., The propagation velocity of world wide sudden commencement of magnetic storms, *J. Geophys. Research*, **63**, 405-408, 1958.
- Dessler, A. J., and E. N. Parker, Hydromagnetic theory of geomagnetic storms, *J. Geophys. Research*, **64**, 2239-2252, 1959.
- Dessler, A. J., W. E. Francis, and E. N. Parker, Geomagnetic storm sudden commencement rise times, *J. Geophys. Research*, **65**, 2715-2719, 1960.
- Eckart, C., Approximate solution of one-dimensional wave equations, *Rev. Mod. Phys.*, **20**, 399-417, 1948.
- Eckart, C., *Hydrodynamics of Oceans and Atmospheres*, Pergamon Press, New York, 1960.
- Eisenhart, L. P., *An Introduction to Differential Geometry*, Princeton University Press, 1947.
- Fejer, J. A., Hydromagnetic wave propagation in the ionosphere, *J. Atmospheric and Terrest. Phys.*, **18**, 135-146, 1960.
- Ferraro, V. C. A., The mean free path in rare ionized gases, *Monthly Notices Roy. Astron. Soc.*, **93**, 416-422, 1933.
- Francis, W. E., and R. Karplus, Hydromagnetic waves in the ionosphere, *J. Geophys. Research*, **65**, 3593-3600, 1960.
- Friedrichs, K. O., and H. Kranzer, Nonlinear wave motion, N.Y.O.-6486, *Notes on Magnetohydrodynamics*, VIII, Inst. of Math. Sci., N. Y. U., 1958.
- Grad, H., Propagation of magnetohydrodynamic waves without radial attenuation in *The Magnetodynamics of Conducting Fluids*, edited by D. Bershader, pp. 37-60, Stanford University Press, 1959.
- Hanson, W. B., and D. D. McKibbin, An ion trap measurement of the ion concentration profile above the F_2 peak, *J. Geophys. Research*, **65**, 1667-1671, 1961.
- Herlofson, N., Magnetic-hydrodynamic waves in a fluid conductor, *Nature*, **165**, 1020-1021, 1950.
- Jacobs, J. A., and K. Sinno, Worldwide characteristics of geomagnetic micropulsations, *Geophys. J.*, **3**, 333-353, 1960.
- Johnson, F. S., The ion distribution above the maximum, *J. Geophys. Research*, **65**, 577-583, 1960.
- Johnson, F. S., The distribution of hydrogen in the telluric hydrogen corona, *Astrophys. J.*, **132**, 701-705, 1961.
- Johnson, F. S., and R. Fish, The telluric hydrogen corona, *Astrophys. J.*, **131**, 502-515, 1960.
- Kanwal, R. P., and C. Truesdell, Electric current and fluid spin created by the passage of a magnetionic wave, *Arch. Rat. Mech. Anal.*, **5**, 423-439, 1960.
- Lighthill, M. J., Studies on magneto-hydrodynamic waves and other anisotropic wave motions, *Phil. Trans. Roy. Soc.*, **252**, 397-430, 1960.
- Newman, P., Optical, electromagnetic and satellite observation of high altitude nuclear detonations, Part I, *J. Geophys. Research*, **64**, 923-932, 1959.
- Piddington, J. H., The transmission of geomagnetic disturbances through the atmosphere and interplanetary space, *Geophys. J.*, **2**, 173-183, 1959.
- Piddington, J. H., Geomagnetic storm theory, *J. Geophys. Research*, **65**, 93-103, 1960a.
- Piddington, J. H., A theory of polar geomagnetic storms, *Geophys. J.*, **3**, 314-332, 1960b.
- Smith, R. L., and R. A. Helliwell, Electron densities to 5 earth radii deduced from nose whistlers, *J. Geophys. Research*, **65**, 2583-2584, 1960.
- Spitzer, L., *The Physics of Fully Ionized Gases*, Interscience Publishers, New York, 1956.
- Storey, L. R. O., An investigation of whistling atmospherics, *Phil. Trans. Roy. Soc.*, **246**, 113-141, 1953.
- Synge, J. L., Elastic waves in anisotropic media, *Math. Phys.*, **35**, 323-334, 1957.
- Troitskaya, V. A., Pulsation of the earth's electromagnetic field with periods of 1 to 15 seconds and their connection with phenomena in the high atmosphere, *J. Geophys. Research*, **66**, 5-11, 1961.
- Tepley, L., Observations of hydromagnetic emissions, *J. Geophys. Research*, **66**, 1651-1658, 1961.
- Townsend, A. A., *The Structure of Turbulent Shear Flow*, Cambridge University Press, 1956.
- Weatherburn, C. E., *Differential Geometry Three Dimensions*, Cambridge University Press, 1955.
- Whitham, G. B., A note on group velocity, *Fluid Mechanics*, **9**, 347-352, 1960.
- Williams, V. L., The simultaneity of sudden commencement of magnetic storms, *J. Geophys. Research*, **65**, 85-92, 1960.

Spatial Distribution of Electrons from Neutron Decay in the Outer Radiation Belt

W. N. HESS AND J. KILLEEN

*Lawrence Radiation Laboratory, University of California
Livermore, California*

Abstract. The spatial distribution of electrons injected into the geomagnetic field from the beta decay of cosmic-ray-produced neutrons leaking out of the atmosphere has been calculated. There is available an IBM 709 code which has been used to calculate the densities of particles trapped by the geomagnetic field for both natural and artificial sources. The spatial distribution of electrons from the neutron decay source in the inner belt has been presented previously. In that calculation losses due to atmospheric scattering were included. In this calculation atmospheric scattering is neglected above 2000 km; but the effect on the mirror-point distribution of slowing down of the electrons is included. These results are compared with some observed spatial distributions of electrons.

Introduction. This paper proposes to show some of the features that the outer zone of the radiation belt would have if it were produced completely by the decay of neutrons leaking out of the atmosphere of the earth. Until recently it had not been supposed that a large fraction of the outer belt was made by neutrons, but this is indeed the case. The flux and energy spectrum of outer belt electrons of $E > 100$ kev is roughly what is expected from neutron decay [Hess, 1960]. Also there are substantial problems [Dessler and Karplus, 1960] with the solar wind and local acceleration mechanisms provided to populate the radiation belt. At energies below 100 kev there are serious discrepancies in experimental measurements. There may or may not be a large low-energy group of electrons in the outer belt. If there is a large group of low-energy electrons, it is not produced by neutron decay but must come from some other source. We have calculated the spatial distribution of the flux of electrons to be expected in the outer radiation belt resulting from neutron decay. Our calculations consider only electrons of $E > 100$ kev for reasons that will be discussed later. This presents nearly all the electrons made by neutron decay. There are only 11 per cent of the calculated equilibrium spectrum of electrons (Hess and Poirier, to be published) having energies below 100 kev.

Several experiments have been carried out in the outer radiation belt that involve counting

electrons by the intermediate step of bremsstrahlung [Fan, Meyer, and Simpson, 1960; Rosen, Farley, and Sonett, 1960; Van Allen and Frank, 1959, 1960; Arnoldy, Hoffman, and Winckler, 1960]. These measurements tend to emphasize higher energy electrons because the detection efficiency in these experiments is a strongly increasing function of energy. Because of this, the information presented in this paper should be useful in interpreting these bremsstrahlung experiments.

To calculate the equilibrium flux of electrons in the outer belt we must understand the source of these electrons and the nature of the losses of the electrons. The only source of electrons that we will consider here is the β decay of neutrons leaking out of the atmosphere of the earth. This source is understood quantitatively [Hess, Canfield, and Lingenfelter, 1961].

There are two loss processes that should be considered. Particles can be lost by multiple small-angle Coulomb scattering changing the pitch angle so that the particles eventually hit the atmosphere and are lost. The other important loss mechanism is the slowing down of electrons by near collisions with thermal electrons.

In the outer radiation belt, where the electrons are moving in the hydrogen exosphere, this slowing down process is very important. At low altitudes, roughly below 2000 km, the electrons move in an atmosphere composed mostly of oxygen and nitrogen. Because of the Z^2 de-

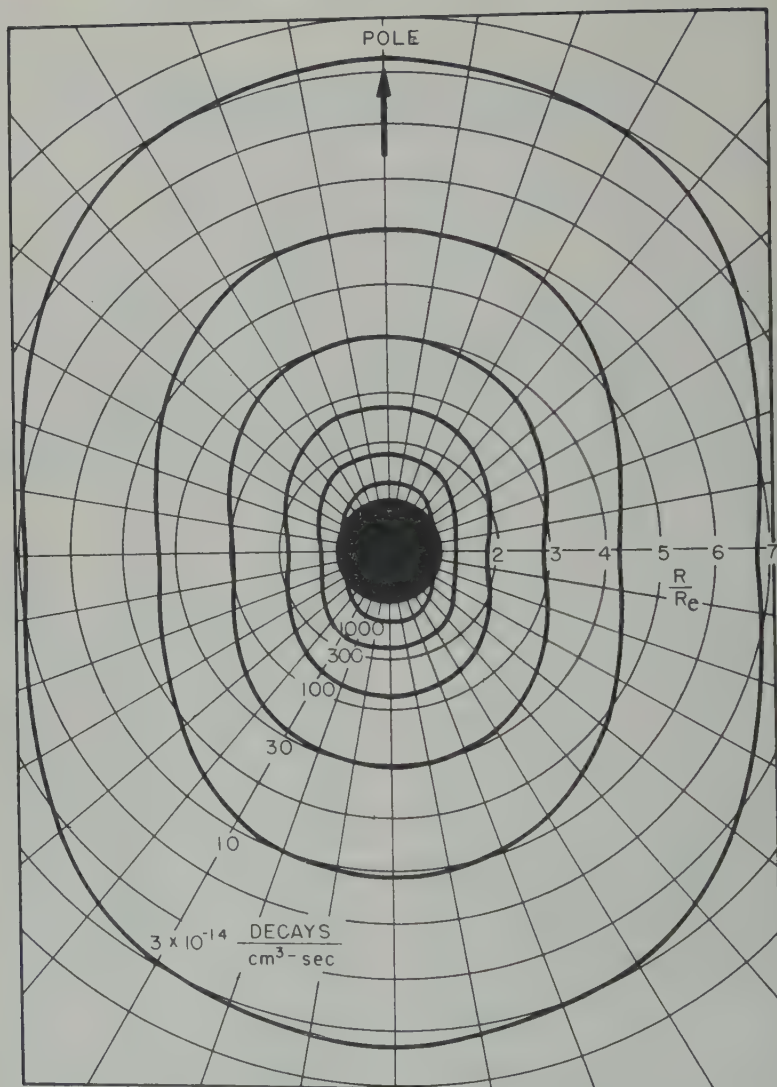


Fig. 1. Neutron decay density in geomagnetic polar coordinates.

pendence of the multiple small-angle Coulomb scattering on nuclear charge, the scattering process dominates the loss mechanisms below 2000 km and we can ignore slowing down at such altitudes.

These two processes have different energy dependencies; therefore one process will be dominant at high energy and the other at low energy. The scattering process goes roughly as E^{-1} . The slowing down process goes roughly as $E^{-0.8}$. Because of these energy dependencies the slowing down process will be more important at high energy and the scattering process more im-

portant at low energy. We have selected 100 km as the energy above which we can consider the scattering is negligible above 2000 km. We consider only the slowing down loss mechanism above 2000 km. The selection of the 100-km energy is somewhat arbitrary and will be discussed later.

Below 2000 km the atmosphere is mostly oxygen and nitrogen and it gets denser rapidly with decreasing altitude. In this region we have used the scattering loss only and have calculated loss rates by using the concept of velocity of lowering of mirror points developed

Christofilos [1959] and by Welch and Whit-
[1959] which applies well in this region.

section mirror-point densities. We wish to
pute the densities of electrons from neutron
y that are trapped by the geomagnetic
The decay density as a function of posi-
in space is known [Hess, Canfield, and
enfelter, 1961]. It is considered as constant
me, so the source density is in electrons/cm³.
We assume that the earth is a sphere and
the geomagnetic field is that of a dipole
ed at the center. We take spherical coordi-
s r, θ, ϕ , where θ is then the geomagnetic
itude measured from the magnetic polar
The source density $n_0(r, \theta)$ is given in this
dinate system, and is shown in Figure 1.
decay source is azimuthally symmetric so
e is no ϕ dependence for n_0 . We wish to
pute the density of mirror points and the
ped-particle density along specific magnetic
lines. The specific lines chosen for this prob-
are shown in Figure 2. These lines are
eterized by the geomagnetic colatitude at
h they meet the earth's surface, i.e., a field
is specified by the value of θ_0 . We must
the source density as a function of position
g the several lines chosen. In order to ex-

press $n_0(r, \theta)$ as $n_0(\theta_0, \theta)$ we use the equation
of the dipole field lines

$$r/R_e = \sin^2 \theta / \sin^2 \theta_0 \quad (1)$$

where R_e is the radius of the earth. The source
density $n_0(\theta_0, \theta)$ is shown in Figure 3.

Given $n_0(\theta_0, \theta)$ we first wish to compute the
density of mirror points, for the line θ_0 , of
particles injected along the line. We assume as
constants of the charged particle motion, v and
 λ , defined by

$$v^2 = v_{\perp}^2 + v_{\parallel}^2 \quad (2)$$

$$\lambda = v_{\perp}^2 / (v^2 B) \quad (3)$$

where B is the magnitude of the field, v_{\perp} and v_{\parallel}
are the components of the particle velocity
perpendicular and parallel to the direction of
the magnetic field. We have from equations
2 and 3 that

$$\mu = v_{\parallel} / v = (1 - \lambda B)^{1/2} \quad (4)$$

$$d\mu = -\frac{1}{2} B (1 - \lambda B)^{-1/2} d\lambda \quad (5)$$

where μ is the cosine of the particles' pitch angle.
The number of particles injected into an element
of volume $dA' dl'$ at a point θ' on a given line

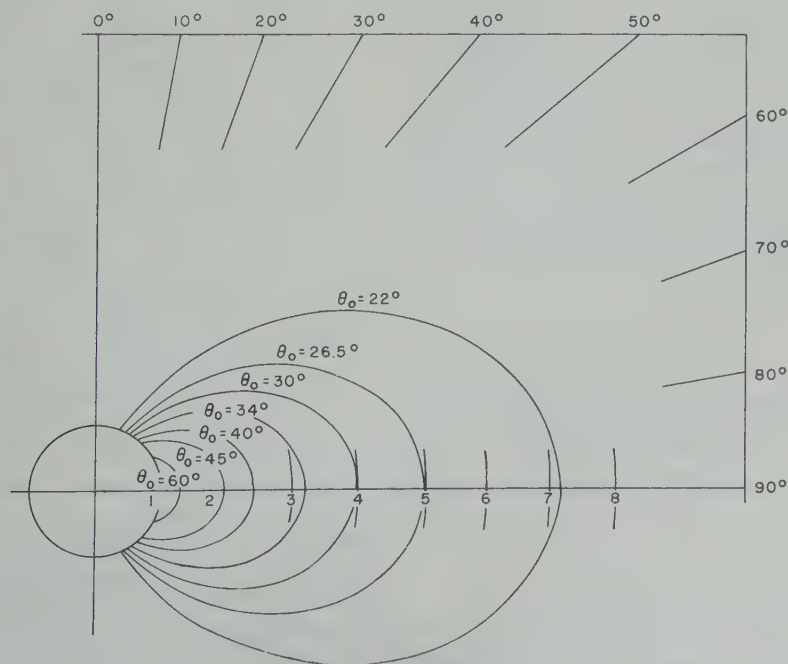


Fig. 2. Magnetic field lines considered in this paper.

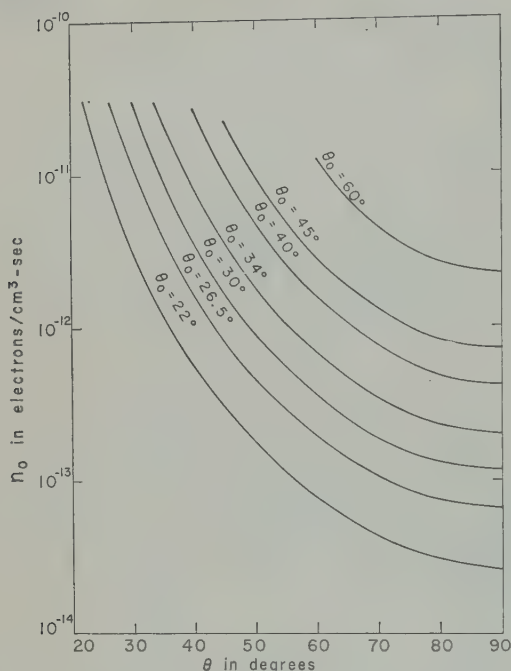


Fig. 3. Neutron decay density along the field lines.

specified by θ_0 , and with velocity directions between μ and $\mu + d\mu$, is given by

$$dv' = n_0(\theta_0, \theta') dA' dl' d\mu$$

Here we are assuming that the distribution of velocities of the injected particles is isotropic. The element dA' is the cross-sectional area of a tube of force, and dl' is the element of arc-length along the line. These particles will be reflected by the increasing magnetic field in a volume $dA dl$ about the point (θ_0, θ_i) . If W denotes the density of turning points, then from dv' we have

$$dW = dv'/dA dl$$

$$= n_0(\theta_0, \theta') dA' dl' d\mu/dA dl$$

At the turning point, $v_i = 0$, so by equation 3

$$\lambda = 1/B(\theta_0, \theta_i) \quad (6)$$

Since dA is a cross section of a tube of force, we have

$$dA = [B(\theta_0, \theta')/B(\theta_0, \theta_i)] dA'$$

$$= \lambda B(\theta_0, \theta') dA' \quad (7)$$

The element dl is determined by the increment of velocity, i.e.,

$$dl = (dl/d\lambda) d\lambda$$

From equation 6 we have

$$d\lambda/dl = -[B(\theta_0, \theta_i)]^{-2} (dB/dl)_{\theta_i}$$

$$= -\lambda^2 (dB/dl)_{\theta_i}$$

Hence, using equations 5-9 we have

$$dW = \frac{1}{2} \lambda (dB/dl)_{\theta_i} \cdot [1 - \lambda B(\theta_0, \theta')]^{-1/2} n_0(\theta_0, \theta')$$

If we sum up the contributions to the turning point density at θ_i from all source points, then

$$W(\theta_0, \theta_i) = \frac{1}{2} (dB/dl)_{\theta_i} [B(\theta_0, \theta_i)]^{-1} \cdot \int_{\pi/2}^{\theta_i} n_0(\theta_0, \theta') \cdot [1 - B(\theta_0, \theta')/B(\theta_0, \theta_i)]^{-1/2} dl' \quad (1)$$

Let $r_0(\theta_0) = r(\theta_0, \pi/2)$, then by equation 1

$$r_0(\theta_0) = R_s / \sin^2 \theta_0 \quad (1)$$

e.g., for $\theta_0 = 45^\circ$, $r_0 = 2R_s$. From equation we also have

$$r/r_0 = \sin^2 \theta \quad (1)$$

Hence

$$dl = [(dr)^2 + r^2(d\theta)^2]^{1/2}$$

$$= r_0(\theta_0) (4 \cos^2 \theta + \sin^2 \theta)^{1/2} \sin \theta d\theta$$

Let $B_0(\theta_0) = B(\theta_0, \pi/2)$, then for the dipole field

$$B(\theta_0, \theta) = B_0(\theta_0) [\tau_0(\theta_0)]^3 r^{-3}$$

$$\cdot (4 \cos^2 \theta + \sin^2 \theta)$$

Using equation 12, we have

$$B(\theta_0, \theta) = B_0(\theta_0) \sin^{-6} \theta \cdot (4 \cos^2 \theta + \sin^2 \theta)^{1/2} \quad (1)$$

Let

$$\beta(\theta) = B(\theta_0, \theta)/B_0(\theta_0)$$

$$= \sin^{-6} \theta (4 \cos^2 \theta + \sin^2 \theta)^{1/2}$$

Let $x = \cos \theta$, $x' = \cos \theta'$, $x_i = \cos \theta_i$, etc., then

$$\beta(x) = (1 - x^2)^{-3} (3x^2 + 1)^{1/2} \quad (1)$$

$$|dl/dx| = r_0(\theta_0) (3x^2 + 1)^{1/2} \quad (1)$$

on 10 then becomes

$$\begin{aligned} (x_i) &= \frac{3}{2}x_i(3x_i^2 + 1)^{-1/2} \\ &- x_i^2)^{-1} + (3x_i^2 + 1)^{-1}] \\ n_0(\theta_0, x') &\{1 - [\beta(x')/\beta(x_i)]\}^{-1/2} \\ &+ 1)^{1/2} dx' \end{aligned} \quad (16)$$

above function is integrated numerically. Although the integrand is singular at the upper limit of integration, the singularity is integrable. Results of the calculation using the neutron source for n_0 are shown in Figure 4.

EQUILIBRIUM MIRROR-POINT DENSITIES

In order to carry out the next step in the calculations we must consider the loss processes. Stated earlier the loss mechanism for outer electrons above 2000 km of $E > 100$ kev slowing down by energy transfer to thermal ions. We can illustrate this by the following example.

Let us consider a magnetic line passing through the equator at 24,000 km geocentric ($\theta_0 = 31^\circ$ latitude). A particle on this line mirroring at 2 km geocentric or 24° latitude is typical outer belt electrons. As this particle slows down from 780 to 100 kev, its pitch angle will change owing to coulomb scattering by a constant root-mean-square scattering angle of 40° . This will spread the mirror points for particles out to slightly past the equator and also spread them down to about 11,000 km. Particles will not be lost from the radiation belt as their mirror points scatter somewhat below 24 km geocentric, so few of these particles are lost to the belt by scattering.

In this example slowing down dominates the loss process completely. Almost half of the scattering here occurs between 200 and 100 kev. As the particle is followed down to 50 kev, the constant rms scattering angle increases to 48° . This will cause a loss of some particles by scattering out of the belt. Based on this kind of argument we have decided that the slowing down mechanism is dominant above 100 kev. This means that the calculations will not be valid at lower energies, and even at above 100 kev there will be some tendency for scattering to change the distributions of particles along a magnetic force. This would seem to be a small

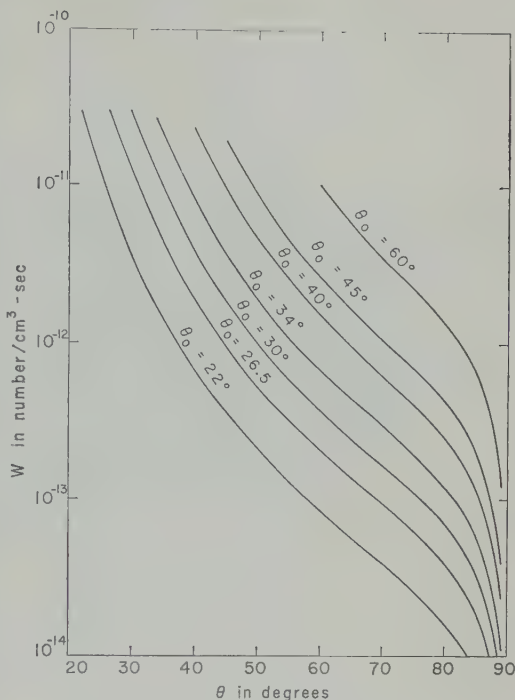


Fig. 4. Injection mirror-point density.

effect, especially since the distribution of particles along a line as produced by scattering [Wentworth, MacDonald, and Singer, 1959] alone is not very different from the one produced by slowing down.

Particles trapped in the earth's magnetic field spend most of their time quite near their mirror points. This is similar to a pendulum, where the period is nearly independent of the amplitude so that the pendulum spends most of its time near its turning point. Consequently, we can say that the density of particles at a particle's mirror point completely determines its lifetime.

Considering the slowing down process we can write $dE = dE/dx) N v d\tau$, where N is the particle density and v is the electron's velocity. The lifetime $\tau(E, \theta_0, \theta_i)$ for a particle of energy E , and mirror point θ_0 , on line θ_0 , is given by

$$\tau(E, \theta_0, \theta_i) = [N(\theta_0, \theta_i)]^{-1} \cdot \int_{E_0}^E \left(\frac{dE}{dx} \right)^{-1} \frac{dE}{v(E)} = \frac{F(E)}{N(\theta_0, \theta_i)} \quad (17)$$

where $N(\theta_0, \theta_i)$ is the particle density at the mirror point θ_i on line θ_0 .

TABLE 1. Values of $\bar{\tau}(\theta_0, \theta_i)$
($\times 10^{10}$ seconds)

θ_i	θ_0					
	45°	40°	34°	30°	26.5°	22°
90°	0.0580	0.0930	0.180	0.350	0.580	0.870
85	0.0552	0.0912	0.175	0.340	0.563	0.866
80	0.0504	0.0869	0.170	0.324	0.537	0.861
75	0.0430	0.0775	0.156	0.292	0.483	
70	0.0341	0.0664	0.138	0.241	0.430	0.791
65	0.0247	0.0547	0.115	0.198	0.358	0.725
60	0.0166	0.0388	0.0878	0.152	0.290	0.588
55	0.0094	0.0232	0.0621	0.111	0.217	0.497
50	0.0046	0.116	0.0391	0.0761	0.147	0.355
45		0.0046	0.0182	0.0467	0.0906	0.231
40			0.0064	0.0212	0.0513	0.134
35			0.0012	0.0053	0.0207	0.0696
30					0.0058	0.0256
25°						0.0043

The average lifetime $\bar{\tau}$ for all particles in the belt is

$$\bar{\tau}(\theta_0, \theta_i) = \int \tau(E, \theta_0, \theta_i) \cdot S(E) dE / \int S(E) dE \quad (18)$$

where $S(E)$ is the equilibrium energy spectrum of electrons in the outer belt. Using a recently calculated energy spectrum (Hess and Poirier, to be published) and values of $N(\theta_0, \theta_i)$ from Johnson [1960], we have calculated values of $\bar{\tau}$ as given in Table 1.

Using these values of $\bar{\tau}$ we go from the injection mirror-point density to the equilibrium mirror-point density by

$$M(\theta_0, \theta_i) = \bar{\tau}(\theta_0, \theta_i) W(\theta_0, \theta_i) \quad (19)$$

The loss mechanism for trapped electrons that we consider below 2000 km is the scattering by the earth's atmosphere. The effect of this scattering is to lower the mirror points of the electrons. As the mirror points move lower, the velocity of lowering increases because of the increasing atmosphere. Let $V(\theta_0, \theta_i)$ denote the velocity of lowering mirror points along the θ_0 line. Consider a specific point θ_i , and consider a tube of force along the θ_0 line for $\theta_i \leq \theta \leq \pi/2$. We subdivide the line as follows: $l_0, l_1, l_2, \dots, l_n$, where l_0 corresponds to $\theta = \pi/2$ and l_n corresponds to $\theta = \theta_i$.

We denote the cross-sectional area of the tube

at the point l_i by ΔA_i and let $\Delta l_i = l_i - l_{i-1}$. Consider the last cell, i.e., the point l_n is at θ_i . The number of mirror points leaving the cell per unit time is given by $M(\theta_0, \theta_n) V(\theta_0, \theta_n) \Delta A_n$. The number of mirror points entering the cell per unit time is given by

$$M(\theta_0, \theta_{n-1}) V(\theta_0, \theta_{n-1}) \Delta A_{n-1} + W(\theta_0, \xi_n) \overline{\Delta A_n} \Delta l_n$$

where

$$\theta_n \leq \xi_n \leq \theta_{n-1} \quad \text{and} \quad \Delta A_n \leq \overline{\Delta A_n} \leq \Delta A_{n-1}$$

Hence

$$\begin{aligned} W(\theta_0, \xi_n) \overline{\Delta A_n} \Delta l_n &= M(\theta_0, \theta_n) V(\theta_0, \theta_n) \Delta A_n \\ &\quad - M(\theta_0, \theta_{n-1}) V(\theta_0, \theta_{n-1}) \Delta A_{n-1} \\ W(\theta_0, \xi_{n-1}) \overline{\Delta A_{n-1}} \Delta l_{n-1} &= M(\theta_0, \theta_{n-1}) V(\theta_0, \theta_{n-1}) \Delta A_{n-1} \\ &\quad - M(\theta_0, \theta_{n-2}) V(\theta_0, \theta_{n-2}) \Delta A_{n-2} \\ &\dots \end{aligned}$$

$$W(\theta_0, \xi_1) \overline{\Delta A_1} \Delta l_1 = M(\theta_0, \theta_1) V(\theta_0, \theta_1) \Delta A_1$$

Adding these equations we have

$$\begin{aligned} \sum_{i=1}^n W(\theta_0, \xi_i) \overline{\Delta A_i} \Delta l_i &= M(\theta_0, \theta_n) V(\theta_0, \theta_n) \Delta A_n \end{aligned}$$

and

$$\overline{\Delta A_i} = [B(\theta_0, \theta_n)/B(\theta_0, \xi_i)] \Delta A_n$$

Hence

$$\begin{aligned} M(\theta_0, \theta_n) V(\theta_0, \theta_n) &= B(\theta_0, \theta_n) \sum_{i=1}^n W(\theta_0, \xi_i) [B(\theta_0, \xi_i)]^{-1} \Delta A_n \end{aligned}$$

Letting $n \rightarrow \infty$ and $\Delta l \rightarrow 0$ we have

$$\begin{aligned} M(\theta_0, \theta_i) &= B(\theta_0, \theta_i) [V(\theta_0, \theta_i)]^{-1} \\ &\quad \cdot \int_{\pi/2}^{\theta_i} W(\theta_0, \theta') [B(\theta_0, \theta')]^{-1} d\theta' \quad (2) \end{aligned}$$

If we use equations 14 and 15, the above expression becomes

$$\begin{aligned} M(\theta_0, x_i) &= r_0(\theta_0) \beta(x_i) [V(\theta_0, x_i)]^{-1} \\ &\quad \cdot \int_0^{x_i} W(\theta_0, x') [1 - x'^2]^3 dx' \quad (2) \end{aligned}$$

In order to compute the equilibrium mirror point density $M(\theta_0, \theta_i)$, we must know the v

of lowering mirror points along each line. Functions $V(\theta_0, \theta_i)$ are computed using the methods of *Christofilos* [1959] and *Welch and Baker* [1959]. They are shown in Figure 5. We see from equation 19 that the equilibrium mirror point density at a particular point θ_i depends on the value of the lifetime at that point. Also from equation 20 we note that the value of M at θ_i is determined by the value of θ_0 . Consequently, we can use one expression for the range of values of θ_i and the other for the different range of values. We use equation 19 for values of θ_i located above 2000 km, and equation 20 for values of θ_i located below 2000 km. The transition for each line is indicated by the end of the curves of Figure 5. The resulting equilibrium mirror point densities are shown in Figure 6.

Trapped-particle densities. We now wish to calculate the trapped-particle density $n(\theta_0, \theta)$. The time taken for a particle to travel from the equator to its reflection point (i.e., the bounce period) is given by

$$T(\theta_0, \theta_i) = \int_{\pi/2}^{\theta_i} \frac{dl}{v_{\parallel}} = \frac{1}{v} \int_{\pi/2}^{\theta_i} [1 - B(\theta_0, \theta)/B(\theta_0, \theta_i)]^{-1/2} dl \quad (22)$$

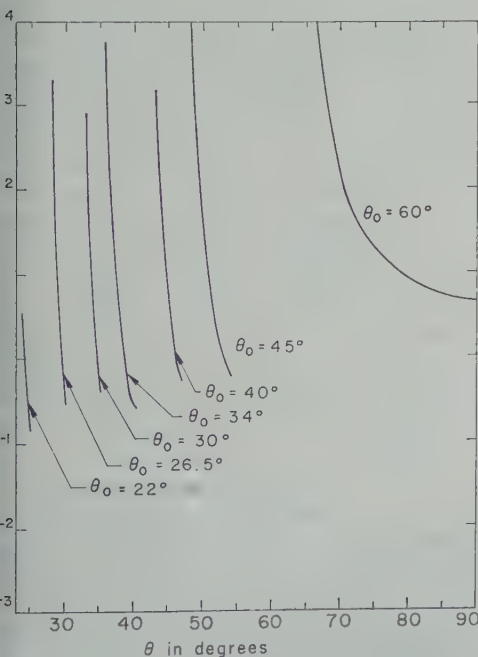


Fig. 5. Velocity of lowering mirror points.

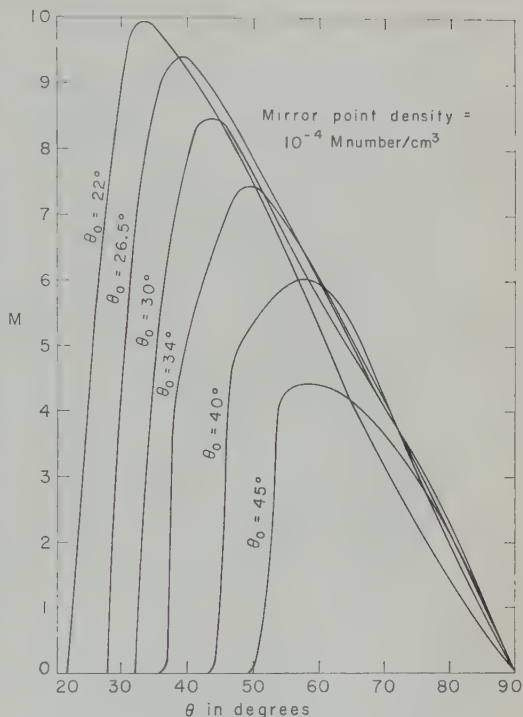


Fig. 6. Equilibrium mirror-point density.

The number of particles with mirror points in the volume $dA_i dl_i$ at the point θ_i is $M(\theta_0, \theta_i) dA_i dl_i$. These particles will spend a time $dl/v_{\parallel}(\theta)$ in the volume element of length dl and cross-section dA at the point θ , i.e., the number of particles with mirror points in $dA_i dl_i$ at θ_i that appear in the volume $dA dl$ at θ is given by

$$[dl/v_{\parallel}(\theta)T(\theta_0, \theta_i)]M(\theta_0, \theta_i) dA_i dl_i$$

The contribution to the number density at θ of particles with mirror points at θ_i is

$$dn = [dl/v_{\parallel}(\theta)T(\theta_0, \theta_i)] \cdot [M(\theta_0, \theta_i) dA_i dl_i/dA dl]$$

We have $dA = [B(\theta_0, \theta_i)/B(\theta_0, \theta)] dA_i$. The contributions to the number density at θ can only come from particles that have mirror points below θ , hence

$$n(\theta_0, \theta) = B(\theta_0, \theta) \int_{\theta}^{\theta_0} M(\theta_0, \theta_i) \cdot [B(\theta_0, \theta_i)vT(\theta_0, \theta_i)]^{-1} \cdot [1 - B(\theta_0, \theta)/B(\theta_0, \theta_i)]^{-1/2} dl_i \quad (23)$$

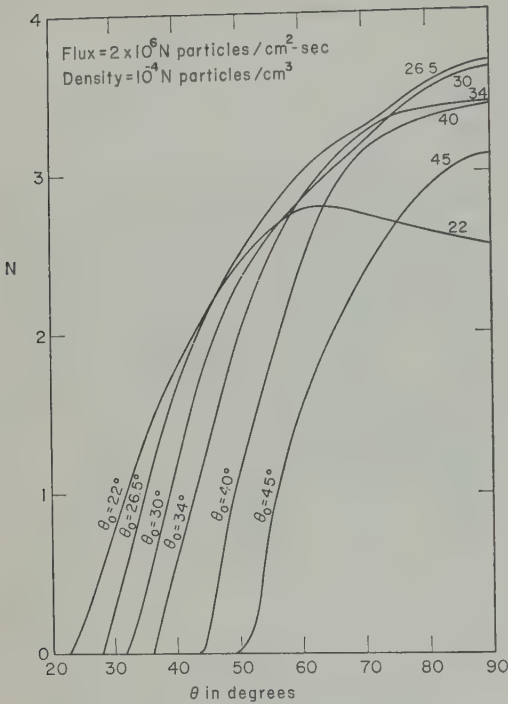


Fig. 7. Trapped particle density and flux.

If we use equations 14 and 15, then

$$T(\theta_0, x_i) = v^{-1}r_0(\theta_0) \int_0^{x_i} [1 - \beta(x)/\beta(x_i)]^{-1/2} \cdot (3x^2 + 1)^{1/2} dx = v^{-1}r_0(\theta_0)F(x_i)$$

Hence equation 23 becomes

$$n(\theta_0, x) = \beta(x) \int_{\cos\theta}^{\cos\theta_0} M(\theta_0, x_i)[\beta(x_i)F(x_i)]^{-1} \cdot [1 - \beta(x)/\beta(x_i)]^{-1/2}[3x_i^2 + 1]^{1/2} dx_i, \quad (24)$$

The results of integrating equation 24 are given in Figure 7. In order to perform the numerical integrations of equations 16 and 24, an IBM 709 code has been written [Killeen and Boer, 1960], which has been used to calculate the densities of particles trapped by the geomagnetic field for both natural and artificial sources. The code was used to predict particle densities from the Argus experiments [Christofilos, 1959], and the spatial distribution of electrons from the neutron decay source in the inner belt [Hess and Killeen, 1960].

Discussion of results. The results of these calculations—the spatial distribution of the flux,

ϕ , of outer-belt electrons produced by neutron decay—are shown in Figure 7. The largest error in the calculations leading to this picture is probably in our knowledge of the density, η , of the hydrogen exosphere. This might be wrong by a factor of 2 or 3 [Johnson, 1960]. Ignoring particles of $E < 100$ kev as we have done probably nowhere produces nearly as large an error as this. The values of the fluxes shown in Figure 7 can be corrected roughly when better exosphere densities become known by using the relation $\phi = \text{const.}/\eta$.

We can compare these flux values and spatial distributions we have calculated with the results of several experiments performed in the outer radiation belt.

The Pioneer III flight [Van Allen and Franz, 1959] measured the flux of particles that count in a GM tube with about 1 g/cm² of shielding. The highest count rate measured on the outgoing flight was 21,400 counts/sec very near the equator at 23,000 km geocentric. On the incoming flight, about 36 hours later, the maximum count rate was 25,600 counts/sec at about 30° λ at about 21,000 km. From Figure 7 we expect the flux to decrease about 30 per cent going from the equator to 30° λ along this magnetic line rather than increase as in the Pioneer III flight.

The Russian Mechta [Vernov, Chudakov, Vakulov, and Logachev, 1959] flight measured relative flux values at different positions along the line $\theta_0 = 30^\circ$. These are shown in the table below.

Altitude, km	ϕ	θ , degrees
400	1	31.0
1800	20	34.5
5600	200	43.1
14000	700	63.5

These data and the Pioneer III data, each normalized at the highest altitude, are shown in Figure 8 compared with our calculation for $\theta_0 = 30^\circ$. The agreement with these data is only moderately good.

The most complete experimental determination of the spatial distribution of electrons along a line of force has recently been completed by Fan, Meyer, and Simpson [1961] based on measurements from the Explorer VI satellite. These results are shown in Figure 9 together with the calculated spatial distributions normalized to the experimental curve at the equator.

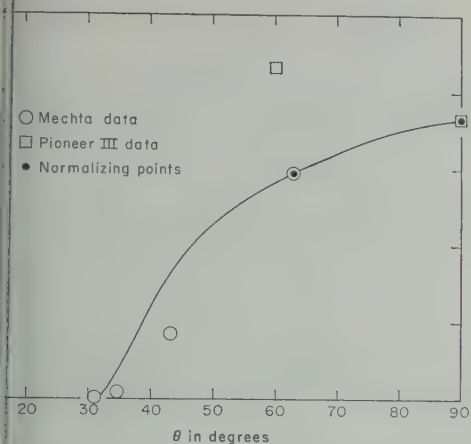


Fig. 8. Comparison of spatial distribution of flux with experiment.

This comparison shows that there are relatively few electrons with mirror points at low latitudes than predicted by the neutron decay hypothesis.

The values of the flux obtained in this calculation within about 30° of the equator, for quite a range of magnetic lines, is about 6×10^6 par/cm²-sec (see Figure 7). Let us compare the calculated flux with the results from several experiments. Experiments which do not involve counting particles by bremsstrahlung and then conversion are the most reliable. The efficiency of counting in bremsstrahlung experiments is difficult to measure and changes rapidly with energy. The Mechta total energy detector (Mechta, Chudakov, Vakulov, and Logachev) counted particles directly, not by bremsstrahlung. This gave about 3×10^{12} ev/cm²-sec energy deposition in the crystal by electrons with $E > 50$ kev at 25° latitude and 20,000 km altitude. From Figure 7 we expect about $6 \times$

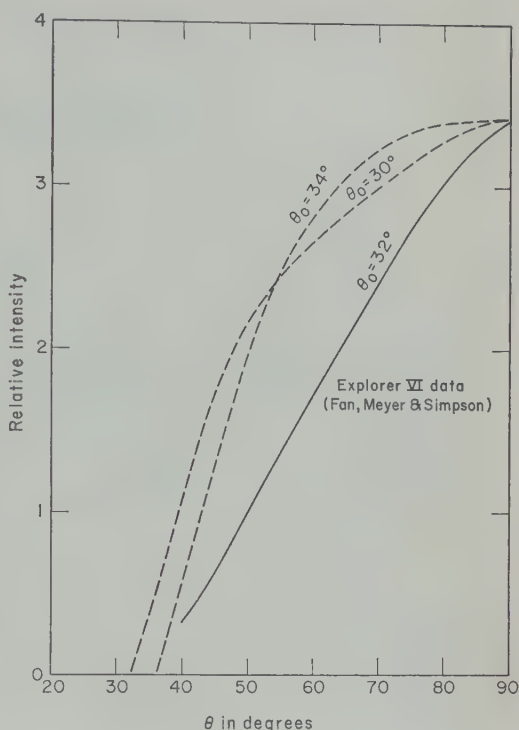


Fig. 9. Comparison of spatial distribution of flux with experiment.

10^6 electrons/cm²-sec at this position ($\theta_0 \approx 30^\circ$, $\theta = 65^\circ$). The average energy of the electrons expected here is about 300 kev (Kellogg, 1960; Hess and Poirier, to be published). This gives the following for the calculated Mechta energy deposition

$$(6 \times 10^6)(0.3 \times 10^6) = 1.8 \times 10^{12} \text{ ev/cm}^2\text{-sec}$$

This agrees relatively well with the experimental value. One other recent experiment that does not depend on bremsstrahlung, by Gringauz,

TABLE 2

Experimenter	Vehicle	ϕ , electrons/cm ² -sec	Energy, kev
Allen, Farley, and Sonnett [1960]	Explorer VI	4×10^7	$E > 500$
Allen and Frank [1959] [1960]	Pioneer III and IV	$\sim 3 \times 10^7$	$E > 200$
Allen and Frank [1959] [1960]	Pioneer III and IV	10^9	$E > 20$
Arnold, Hoffman, and Winckler [1960]	Explorer VI	1.2×10^{11}	$E > 30$

Allen and Frank quote 10^{11} electrons/cm²-sec for Pioneer IV which was during a magnetically disturbed time. Pioneer III saw about two orders of magnitude less flux than Pioneer IV, and this during relatively quiet time, so we have used this value.

Kurt, Moroz, and Shklovskii [1960], used an ion trap to collect particles. They found $\phi \leq 2 \times 10^7$ electrons/cm²-sec of $E > 200$ ev for the maximum flux in the outer radiation belt. This is three times larger than our calculated value but within our error in knowledge of exospheric density, so it does not disagree badly. Even flux values from several bremsstrahlung experiments in Table 2 are not much larger than this in some cases. These values are for magnetically quiet times. These experiments have been discussed by Hess [1960] and Dessler [1960]. The experiments with higher energy limits seem to have values moderately close to our calculated value. For the two lower threshold experiments the values are considerably higher than our calculated values. They also seem to disagree with the Mechta data [Vernov, Chudakov, Vakulov, and Logachev, 1959] and with Gringauz, Kurt, Moroz, and Shklovskii [1960]. The important fact here is that, at least above 50 kev, a significant fraction of the electrons in the outer belt do seem to come from neutron decay. We will have to wait for more detailed experiments to find out whether the fraction is 10 or 100 per cent.

Acknowledgments. We wish to thank Mr. Garrett Boer and Mrs. Julia Kleinecke for valuable assistance in performing the calculations. This work was done under the auspices of the U. S. Atomic Energy Commission.

REFERENCES

- Arnoldy, R., R. Hoffman, and J. R. Winckler, Observations of the Van Allen radiation regions during August and September 1959, *J. Geophys. Research*, **65**, 1361-1376, 1960.
- Christofilos, N. C., The Argus experiment, *J. Geophys. Research*, **64**, 869-875, 1959.
- Dessler, A. J., and R. Karplus, Some properties of the Van Allen Radiation, *Phys. Rev. Letters*, **4**, 271-274, 1960.
- Dessler, A. J., Discussion of a paper by R. L. Arnoldy, R. A. Hoffman, and J. R. Winckler, *J. Geophys. Research*, **65**, 3487-3490, 1960.
- Fan, C. Y., P. Meyer, and J. A. Simpson, Trapped and cosmic radiation measurements from Explorer VI, First International Space Science Symposium, North Holland Publishing Company, Amsterdam, 1960.
- Fan, C. Y., P. Meyer, and J. A. Simpson, The equatorial pitch angle distributions of electrons in the outer radiation belt, Second International Meeting of COSPAR, Florence, Italy, April 1961. (This is also University of Chicago report EFINS-61-20.)
- Gringauz, K. I., V. G. Kurt, V. I. Moroz, and I. Shklovskii, Ionized gas and fast electrons in the vicinity of earth and in interplanetary space, *Doklady Akad. Nauk SSSR*, **132**, 1062-1064, 1960.
- Hess, W. N., The Radiation belt produced by neutrons leaking out of the atmosphere of the earth, *J. Geophys. Research*, **65**, 3107-3115, 1960.
- Hess, W. N., E. Canfield, and R. E. Lingenfelser, Cosmic-ray neutron demography, *J. Geophys. Research*, **66**, 665-677 (1961).
- Hess, W. N., and J. Killeen, Densities of electrons from neutron decay trapped by the geomagnetic field, *Bull. Am. Phys. Soc.*, (2) **5**, 260, 1960.
- Johnson, F. S., The ion distribution above the F_2 maximum, *J. Geophys. Research*, **65**, 557-564, 1960.
- Kellogg, P. J., Electrons of the Van Allen Radiation, *J. Geophys. Research*, **65**, 2705-2715, 1960.
- Killeen, J., and G. Boer, Densities of particles trapped by the geomagnetic field, *Bull. Am. Phys. Soc.* (2) **5**, 312, 1960.
- Rosen, A., T. A. Farley, and C. P. Sonett, Solar radiation measurements of Explorer VI earth satellite, First International Space Science Symposium, North Holland Publishing Company, Amsterdam, 1960.
- Van Allen, J. A., and L. A. Frank, Radiation around the earth to a radial distance of 107,400 kilometers, *Nature*, **183**, 430-434, 1959.
- Van Allen, J. A., and L. A. Frank, Radiation measurements to 658,300 kilometers with Pioneer 1, *Nature*, **184**, 219, 1960.
- Vernov, S. N., A. E. Chudakov, P. V. Vakulov, and Yu. I. Logachev, A study of terrestrial corpuscular radiation and cosmic rays during the flight of a cosmic rocket, *Doklady Akad. Nauk SSSR*, **125**, 304-307, 1959.
- Welch, J. A., and W. A. Whitaker, Theory of geomagnetically trapped electrons from an artificial source, *J. Geophys. Research*, **64**, 909-915, 1959.
- Wentworth, R. C., W. M. MacDonald, and S. Singer, Lifetimes of trapped radiation-belt particles determined by Coulomb scattering, *Planet. Fluids*, **2**, 499-509, 1959.

(Manuscript received July 17, 1961; revised August 28, 1961.)

Coordinates for Mapping the Distribution of Magnetically Trapped Particles

CARL E. McILWAIN

*Department of Physics and Astronomy
State University of Iowa
Iowa City, Iowa*

Abstract. Dipole representations of the earth's magnetic field have been found to have sufficient accuracy for the study of magnetically trapped particles. A coordinate system consisting of the magnitude of the magnetic field B and the integral invariant I has been found to organize adequately measurements made at different geographic locations. The present paper shows that a parameter $L = f(B, I)$ can be defined that retains most of the desirable properties of I and that has the additional property of organizing measurements along lines of force. Since the parameter L is the analog of a physical distance in a dipole field (the equatorial radius of a magnetic shell), it is usually found to present fewer conceptual difficulties than the integral invariant I .

Introduction. Considerable difficulties were encountered in the early attempts to map the field intensities of trapped particles. The difficulties were chiefly due to the high spatial gradients in the intensities. At low altitudes in the inner zone, for example, a change in the distance from the center of the earth of 3 per cent results in an intensity change of over a factor of 10. The shape of the earth's field deviates more than 3 per cent from any possible dipole representation, and therefore attempts to use a dipole model for a coordinate system can be expected to produce rather chaotic results.

The present paper describes a coordinate system that not only takes account of the non-spherical character of the field but can also be used to organize measurements along lines of force as necessary for most theoretical studies.

Adiabatic invariants. The use of adiabatic invariants in describing the motion of a particle in a magnetic field has been discussed by North and Teller [1960]. In the present paper, the longitudinal (or longitudinal) invariant at a point A is defined as

$$I = \oint_A^{A'} (1 - B_i/B)^{1/2} ds$$

where ds is the differential path length along the line of force connecting the point A with its mirror point A' , B_i is the magnitude of the magnetic field along the line of force, B is the magnitude of the magnetic field at the point A ,

and where the integral is to be taken along the line of force between the conjugate points. Defined in this manner, I can be considered to be a scalar field that has a definite value at each point in space and does not require reference to the motion of trapped particles. If there is no electric field, this definition is equivalent to

$$I = P^{-1} \oint_A^{A'} P_{\parallel} ds$$

where P is the total momentum of a particle that mirrors at the point in space A , and where P_{\parallel} is the component of momentum along the line of force. If we assume that all three adiabatic invariants are conserved for a given particle, then the values of I are the same at every point in space at which a given particle mirrors. If the energy and mirror point distributions of a set of trapped particles in time equilibrium do not change importantly during the time the slowest particles drift once around the earth (for example, the high energy protons in the inner zone), then the directional intensity perpendicular to the line of force will be the same at all points in space having the same values of B and I .

Magnetic shells. The points in space that have the same value of B and I form a ring in each hemisphere. A particle mirroring at this B and I will remain upon the surface (or 'shell') described by the lines of force that connect these rings.

In general, two particles that initially mirror

at different values of B along a particular line of force will not drift in longitude to the same lines of force. This means that the shells described by the trajectories of the two particles do not coincide. It will be shown however, that in the earth's magnetic field, this effect is relatively small. One important consequence of this fact is that the omnidirectional intensity as well as the directional intensity is constant along the loci of constant B and I .

Definition of the magnetic shell parameter L . The fact that all particles that drift through a given line of force will remain on approximately the same shell throughout their motion leads immediately to the desirability of finding a method of labeling all points in space with a number that is unique for each shell.

Each locus of constant B and I is a line that lies on a fixed shell. Any parameter that is a function of only B and I will, therefore, have the correct longitude dependence. The problem of defining a magnetic shell parameter is thus reduced to finding a function $f(B,I)$ that is constant along lines of force. Such a function could be found empirically by examining the functions $I_s = g_s(B)$ along lines of force calculated with a representation of the earth's field. An empirical fit is not necessary, however, because the average of the functions $g_s(B)$ around a shell is well represented by the function obtained for a dipole field.

The integral invariant in a dipole field at a magnetic latitude λ on a line of force having an equatorial radial distance of R_0 is given by

$$I = 2R_0 \int_0^Y \left\{ 1 - \left[\frac{(1 + 3Y_a^2)}{(1 + 3Y^2)} \right]^{1/2} \cdot \left[\frac{(1 - Y^2)}{(1 - Y_a^2)} \right]^{3/2} \right\} (1 + 3Y_a^2)^{1/2} dY_a \tag{1}$$

or

$$I = R_0 h_1(\lambda) \tag{2}$$

where $Y = \sin \lambda$. Now in a dipole field

$$B = \frac{M}{R_0^3 \cos^6 \lambda} (1 + 3 \sin^2 \lambda)^{1/2} \tag{3}$$

where M is the dipole magnetic moment. By (3) we see that

$$R_0^3 B/M = h_2(\lambda) \tag{4}$$

Equations (2) and (4) give

$$\begin{aligned} I^3 B/M &= R_0^3 B/M h_1^3(\lambda) \\ &= R_0^3 B/M h_3(R_0^3 B/M) = h_4(R_0^3 B/M) \end{aligned}$$

or

$$R_0^3 B/M = F(I^3 B/M) \tag{5}$$

The magnetic shell parameter L is now defined for a point in the earth's magnetic field by the equation

$$L^3 B/M = F(I^3 B/M) \tag{6}$$

where I and B are to be calculated for the point with a representation of the earth's field, M is the dipole moment of the earth,¹ and F is the function in equation 5 which is calculated with a dipole field.

A set of values for the function F is given in Table 1. For accurate computation of L , the following method can be used:

Let

$$\ln \left(\frac{L^3 B}{M} - 1 \right) = \sum_{n=0}^{n=6} a_n X^n \tag{7}$$

where $X = \ln(I^3 B/M)$. Sets of the coefficients a_n for different ranges of X are given in Table

TABLE 1. Values for Function F

$I^3 B/M$	$L^3 B/M$
2.37529×10^{-15}	1.00002
1.72697×10^{-13}	1.00016
3.70298×10^{-11}	1.00045
2.37086×10^{-9}	1.00180
2.70390×10^{-8}	1.00406
1.52215×10^{-7}	1.00722
5.82114×10^{-7}	1.01131
1.74352×10^{-6}	1.01632
4.41257×10^{-6}	1.02228
9.87364×10^{-6}	1.02919
2.01130×10^{-5}	1.03707
3.80506×10^{-5}	1.04596
6.78113×10^{-5}	1.05585
1.15049×10^{-4}	1.06680
1.87314×10^{-4}	1.07881
2.94482×10^{-4}	1.09194
4.49250×10^{-4}	1.10620
6.67702×10^{-4}	1.12165
9.69961×10^{-4}	1.13832
1.38095×10^{-3}	1.15626

¹ The value $M = 8.06 \times 10^{25}$ gauss cm³ 0.311653 gauss R_e^3 ($R_e = 6371.2$ km) has been found to be satisfactory.

TABLE 1. Continued

$I^3 B/M$	$L^3 B/M$
1.93121×10^{-3}	1.17553
2.65795×10^{-3}	1.19618
3.60595×10^{-3}	1.21827
4.82943×10^{-3}	1.24187
6.39281×10^{-3}	1.26706
8.37277×10^{-3}	1.29392
1.08602×10^{-2}	1.32252
1.39624×10^{-2}	1.35298
1.78056×10^{-2}	1.38538
2.25380×10^{-2}	1.41985
2.83332×10^{-2}	1.45650
3.53938×10^{-2}	1.49547
4.39567×10^{-2}	1.53691
5.42973×10^{-2}	1.58097
6.67358×10^{-2}	1.62782
8.16455×10^{-2}	1.67767
9.94598×10^{-2}	1.73070
1.20681×10^{-1}	1.78716
1.45893×10^{-1}	1.84729
1.75772×10^{-1}	1.91135
2.11103×10^{-1}	1.97966
2.52798×10^{-1}	2.05255
3.01912×10^{-1}	2.13036
3.59672×10^{-1}	2.21352
4.27504×10^{-1}	2.30245
5.07058×10^{-1}	2.39766
6.00263×10^{-1}	2.49967
7.09352×10^{-1}	2.60911
8.36934×10^{-1}	2.72664
9.86054×10^{-1}	2.85300
1.16025	2.98905
1.36367	3.13571
1.60116	3.29403
1.87843	3.46520
2.20216	3.65056
2.58020	3.85161
3.02183	4.07005
3.53799	4.30783
4.14164	4.56716
4.84816	4.85057
5.67587	5.16095
6.64657	5.50164
7.78642	5.87648
9.12680	6.28994
1.07055×10	6.74720
1.25680×10	7.25432
1.47699×10	7.81840
1.73787×10	8.44781
2.04766×10	9.15245
2.41655×10	9.94413
2.85705×10	1.08369×10
3.38471×10	1.18478×10
4.01899×10	1.29973×10
4.78427×10	1.43104×10
5.71146×10	1.58179×10
6.83985×10	1.75576×10
8.21991×10	1.95767×10
9.91695×10	2.19347×10

TABLE 1. Continued

$I^3 B/M$	$L^3 B/M$
1.20162×10^2	2.47069×10
1.46301×10^2	2.79896×10
1.79085×10^2	3.19079×10
2.20532×10^2	3.66255×10
2.45179×10^2	3.93495×10
2.73402×10^2	4.23601×10
3.05069×10^2	4.56967×10
3.41513×10^2	4.94049×10
3.82642×10^2	5.35387×10
4.30241×10^2	5.81615×10
4.84311×10^2	6.33486×10
5.47285×10^2	6.91896×10
6.19343×10^2	7.57921×10
7.03884×10^2	8.32855×10
8.01427×10^2	9.18267×10
9.16824×10^2	1.01607×10^2
1.05125×10^3	1.12862×10^2
1.21182×10^3	1.25883×10^2
1.40095×10^3	1.41032×10^2
1.62942×10^3	1.58767×10^2
1.90201×10^3	1.79669×10^2
2.23574×10^3	2.04483×10^2
2.64005×10^3	2.34175×10^2
3.14300×10^3	2.70014×10^2
3.76361×10^3	3.13689×10^2
4.55081×10^3	3.67477×10^2
5.54426×10^3	4.34504×10^2
6.83524×10^3	5.19133×10^2
8.51087×10^3	6.27578×10^2
1.07567×10^4	7.68892×10^2
1.37789×10^4	9.56603×10^2
1.79974×10^4	1.21153×10^3
2.39555×10^4	1.56683×10^3
3.27489×10^4	2.07743×10^3
4.60402×10^4	2.83877×10^3
6.73169×10^4	4.02645×10^3
1.02942×10^5	5.98740×10^3
1.67841×10^5	9.47159×10^3
2.96984×10^5	1.63048×10^4
5.95512×10^5	3.17247×10^4
1.44453×10^6	7.49153×10^4
5.00559×10^6	2.51887×10^5
1.47748×10^7	7.32706×10^5
4.09266×10^7	2.00752×10^6
1.91226×10^8	9.28013×10^6
6.50554×10^8	3.12969×10^7
5.20779×10^9	2.50188×10^8

2. This method introduces an error in L of less than 0.3 per cent for $-\infty < X < \infty$ and less than 0.03 per cent for $X < 10$.

Significance of L . In general, L should be regarded as a parameter that retains most of the useful properties of I exactly and that is also approximately constant along lines of force. To gain some feeling for the meaning of L , it is

TABLE 2. Coefficients a_n for Different Ranges of X

	$X < -16$	$-16 < X < 0$	$0 < X < 8$	$8 < X < 21$	$X > 21$
a_0	0.294	0.62290	0.62291	1.0824	-3.04
a_1	0.330	0.43351	0.43416	0.20395	1.00
a_2	0	1.4495×10^{-2}	1.3680×10^{-2}	5.4145×10^{-2}	0
a_3	0	1.2154×10^{-3}	1.4784×10^{-3}	-9.3218×10^{-4}	0
a_4	0	5.9474×10^{-5}	1.2413×10^{-5}	-5.6831×10^{-5}	0
a_5	0	1.5367×10^{-6}	-8.1278×10^{-6}	2.7879×10^{-6}	0
a_6	0	1.5843×10^{-8}	1.4604×10^{-7}	-3.4751×10^{-8}	0

useful to remember that, for a pure dipole field, a magnetic shell is labeled by an L equal to the shell's equatorial radius.
For most purposes, B and L should be used

as the principal spatial coordinates. For studies involving intuition, it is sometimes desirable to use a coordinate system having a closer resemblance to the actual physical geometry.

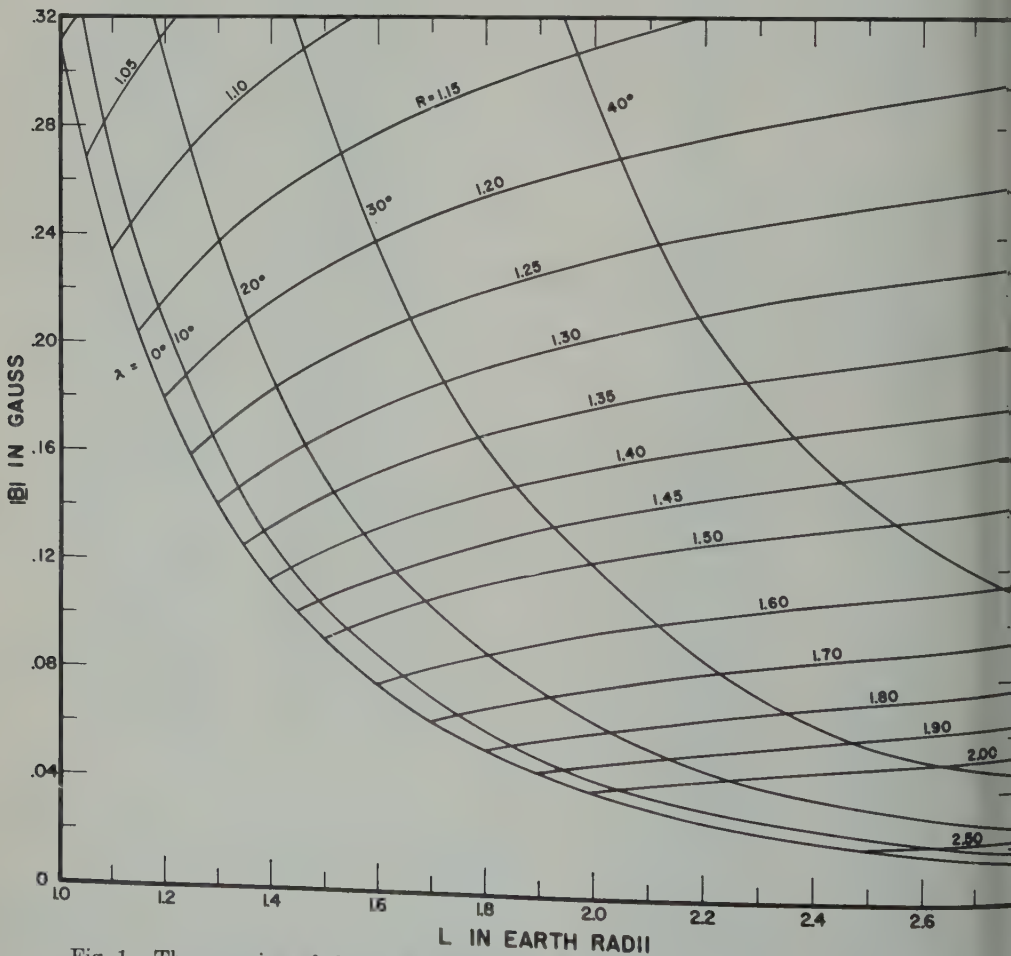
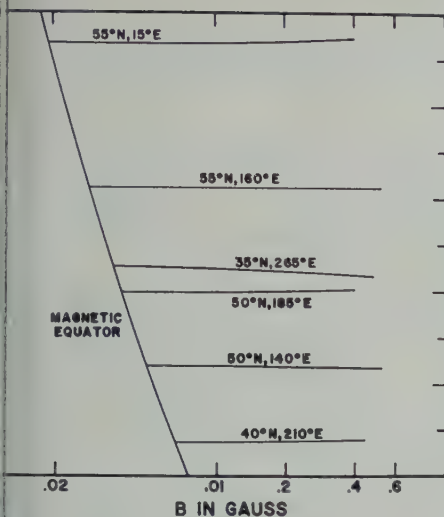


Fig. 1. The mapping of the polar coordinates R and λ on to the B, L plane according to the transformation

$$B = \frac{M}{R^3} \left(4 - \frac{3R}{L} \right)^{1/2}$$
$$R = L \cos^2 \lambda$$



The variation in L along six lines of force. Several of the worst cases are included.

system of this kind can be obtained by transforming B and L to polar coordinates using the dipole relations

$$B = \frac{M}{R^3} \left(4 - \frac{3R}{L} \right)^{1/2}, \quad R = L \cos^2 \lambda \quad (8)$$

This relationship is illustrated in Figure 1.

A great deal of caution must be exercised when using the coordinates R and λ obtained in this manner, because although the irregular characteristics of the magnetic field will in effect be removed, geographic coordinates will transform in an irregular-longitude dependent manner.

Variation of L along lines of force. Tables of I versus B along 1400 different lines of force were kindly supplied to the author by D. C. Jensen, R. W. Murray, and J. A. Welch, Jr. These tables were computed with the 512 term

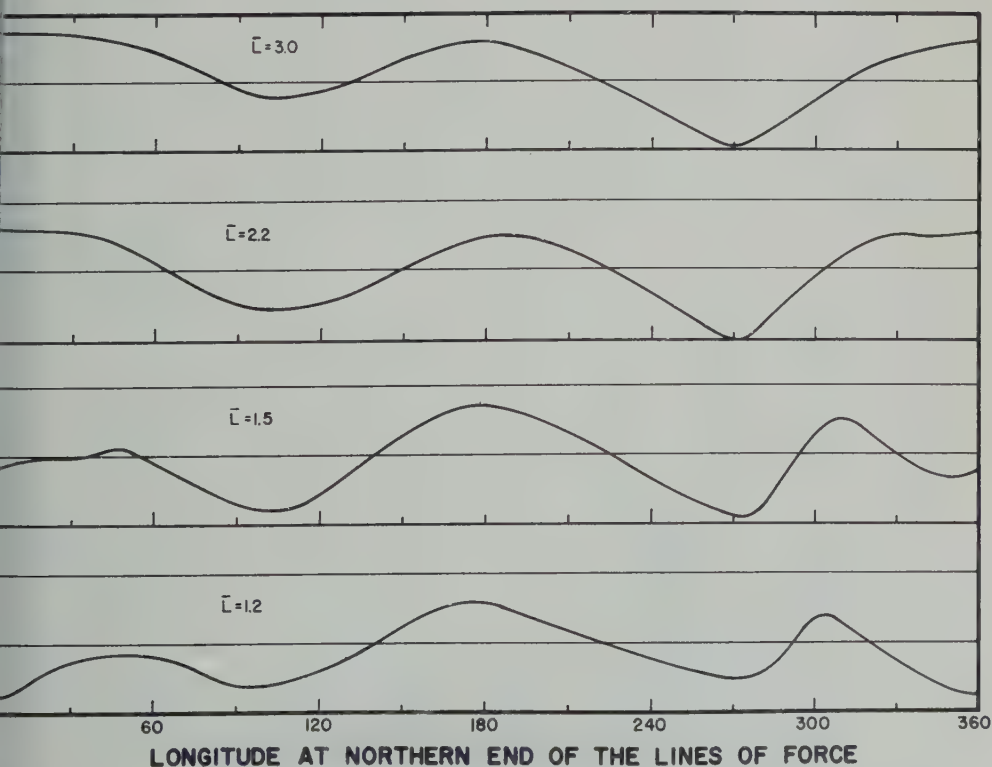


Fig. 3. The maximum deviation in L along lines forming magnetic shells labeled by $L = 1.2, 1.5, 2.2$, and 3.0 earth radii plotted versus the geographic longitude at the northern ends of the lines. A positive deviation means that L increased toward higher values of B .

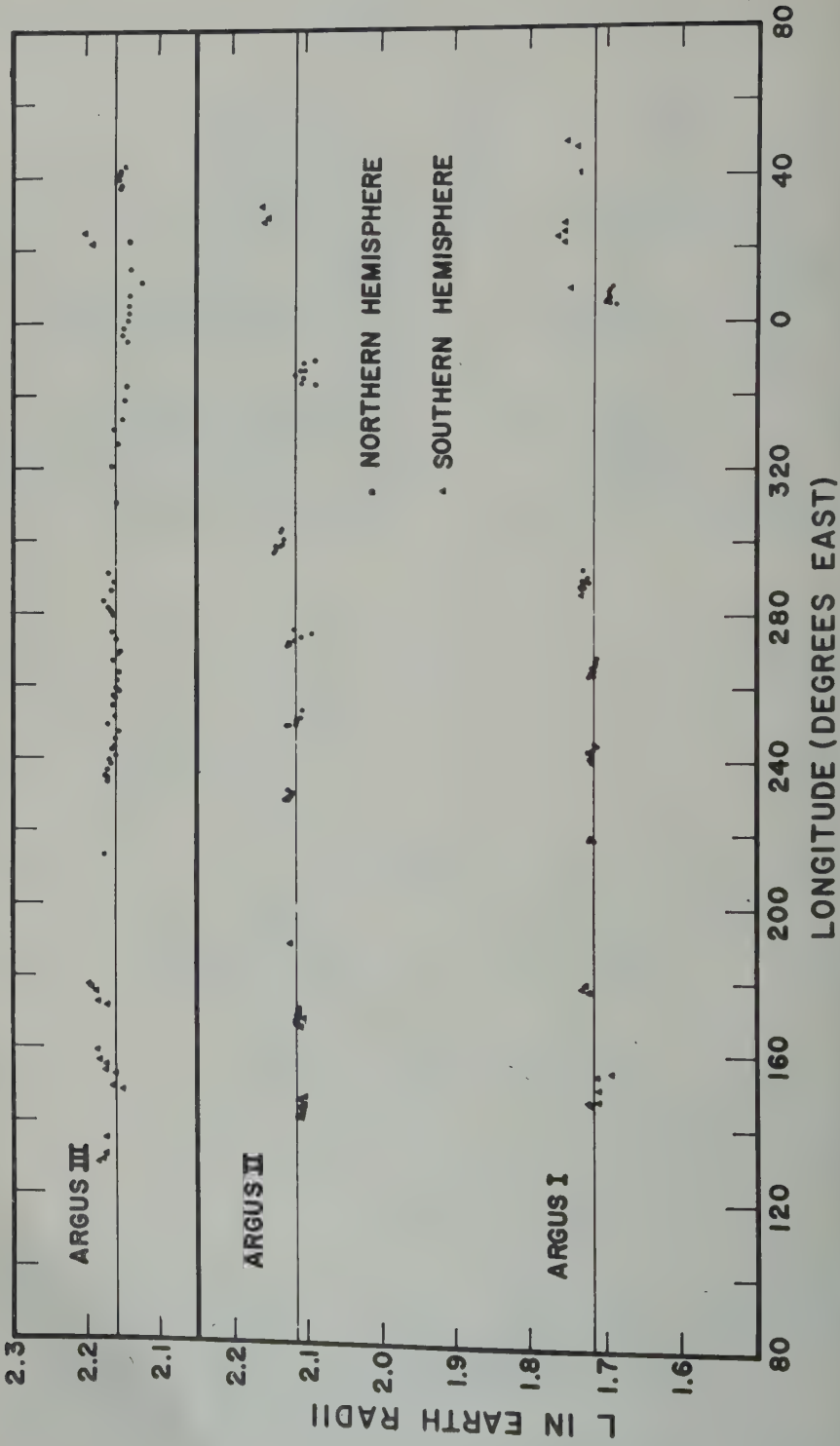


Fig. 4. The L values for the observed intersections of the Explorer IV trajectory with the three Argus shells are plotted versus geographic longitude.

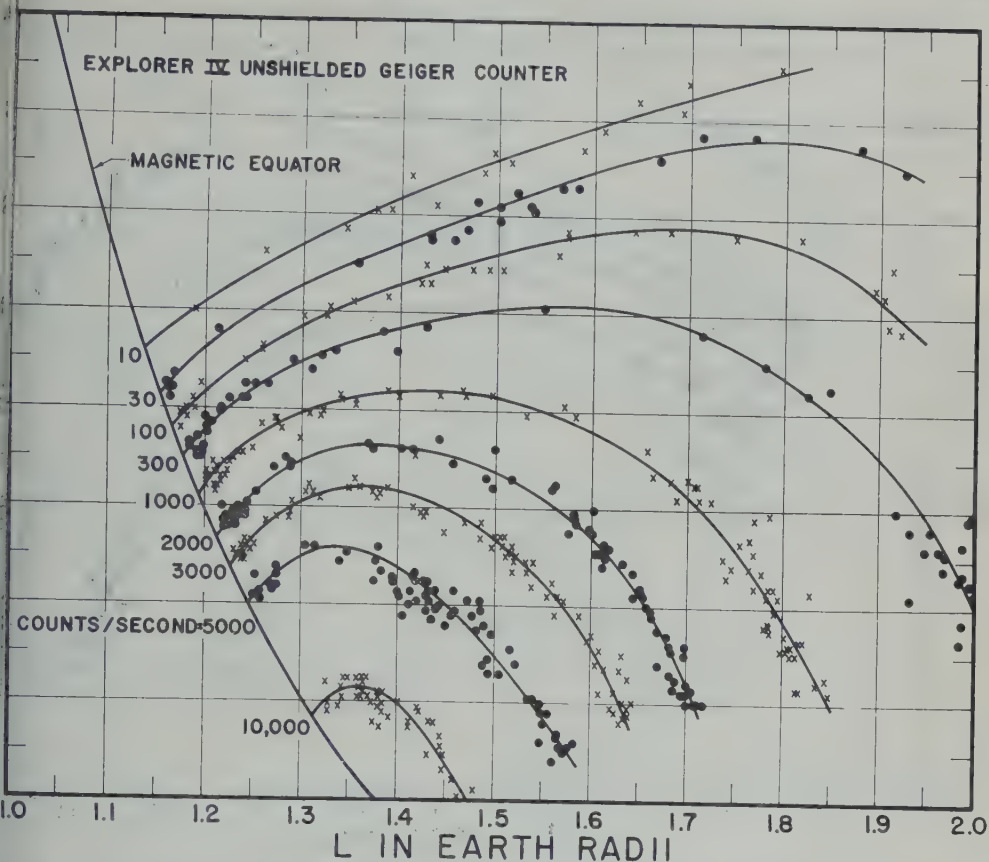


Fig. 5. Contours of constant true counting rate of the unshielded Geiger counter in Explorer IV. The points shown correspond to data obtained over a wide range of geographic latitude and longitude.

al harmonic expansion ($n = 24$; $m = 17$) of the earth's field obtained by Jensen and Scher [1960] and from the 1955 surface field. The L values obtained from these sets of data were found to vary less than 1 per cent from most of the lines of force. Figure 2 shows the variation in L along six lines of force and includes several of the worst cases. The variations along different lines of force were found to be systematic in both latitude and longitude. This is illustrated in Figure 3 where the deviation of L , along lines forming the magnetic field, is plotted versus the geographic longitude at the northern ends of the lines. A positive deviation means that L increased toward higher latitudes.

A check upon the variation of L along lines of force tends to be independent of

the precision of the magnetic field analysis. An analysis of the earth's field based upon improved knowledge of the surface field would probably have nondipole terms of similar magnitude and therefore would be expected to give similar variations in L along lines of force. It should be noted, however, that, in the present analysis, the contributions of external current systems have been assumed to be zero.

The Argus shells. Well-defined shells of electrons injected by the three high altitude nuclear detonations Argus I, II, and III were detected by Satellite 1958 ϵ (Explorer IV) [Van Allen, McIlwain, and Ludwig, 1959b]. Each of the observed intersections of the satellite trajectory with an Argus shell should have the same value of L . Figure 4 shows the L values obtained for the observed intersections as a function of geographic longitude. The average L values for the

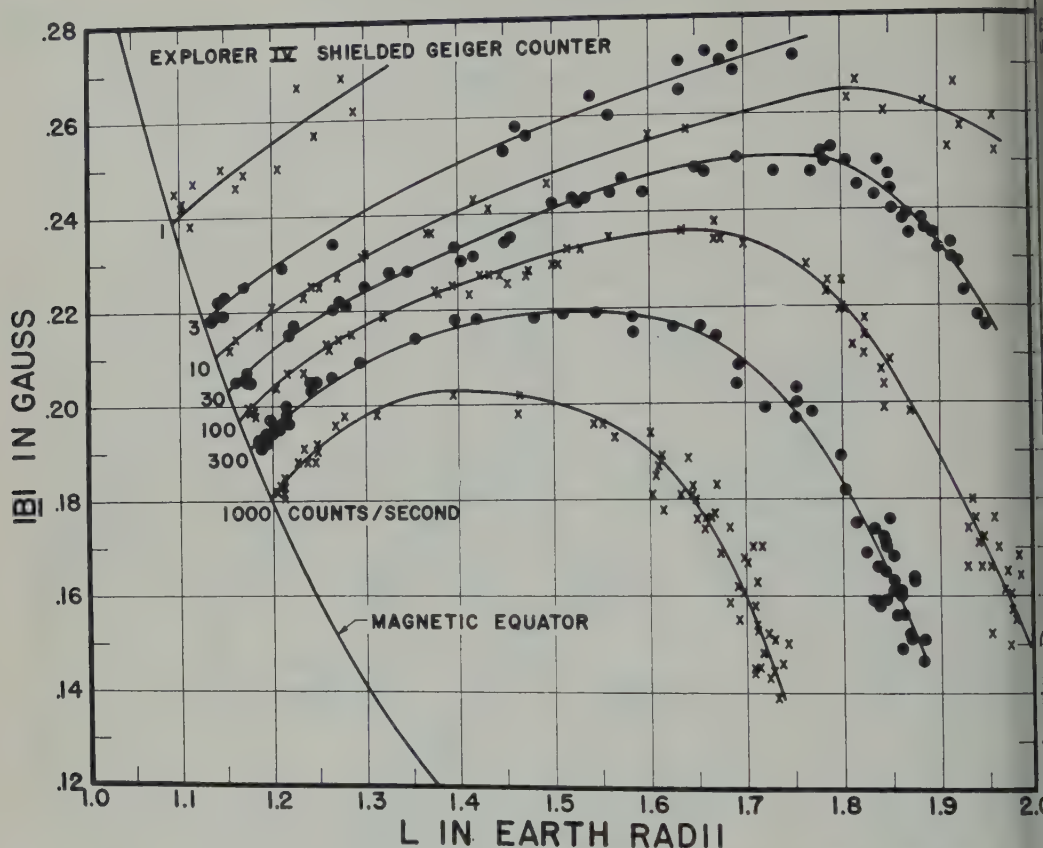


Fig. 6. Contours of constant true counting rate of the shielded Geiger counter in Explorer IV.

shells are 1.715, 2.115, and 2.16 earth radii for Argus I, II, and III, respectively. In addition to the scatter due to measurement inaccuracies, systematic variations with longitude of approximately ± 1.5 per cent can be discerned. By comparison, Pennington [1961] has found that the trace of the data points in the northern hemisphere for Argus III projected upon the surface of the earth deviates more than $\pm 3.5^\circ$ from the earth trace of the best possible offset dipole shell. A latitude error of $\pm 3.5^\circ$ for the Argus III shell corresponds to an error in L of about ± 13 per cent.

The range of B for the data points on each shell is not large. The deviations of the L values in Figure 4 are, therefore, almost entirely due to inaccuracies in the surface field upon which the magnetic field representation is based. It is interesting to note that the deviations in the L values are largest in the regions where the discrepancies between the various available maps

of the surface field are also large. No definite time dependence can be discerned in the data shown in Figure 4. An upper limit on the rate of change in L is 0.001 earth radii per day.

Proton intensities in the inner zone. The most complete set of data pertaining to the spatial distribution of particles in the inner zone presently available was obtained by the Geiger counters on Satellite 1958 ϵ [Van Allen, McIlwain, and Ludwig, 1959a]. The results of a preliminary analysis of the Geiger tube data are shown in Figures 5 and 6. The geometric factors at the proton thresholds are approximately 0.54 cm² and 31 Mev for the unshielded counter and 0.05 cm² and 43 Mev for the shielded counter. Comparison with the scintillation counter data indicates that most of the Geiger tube counting rates were probably due to penetrating protons rather than bremsstrahlung from the very much higher fluxes of nonpenetrating electrons.

The data points shown in Figures 5 and

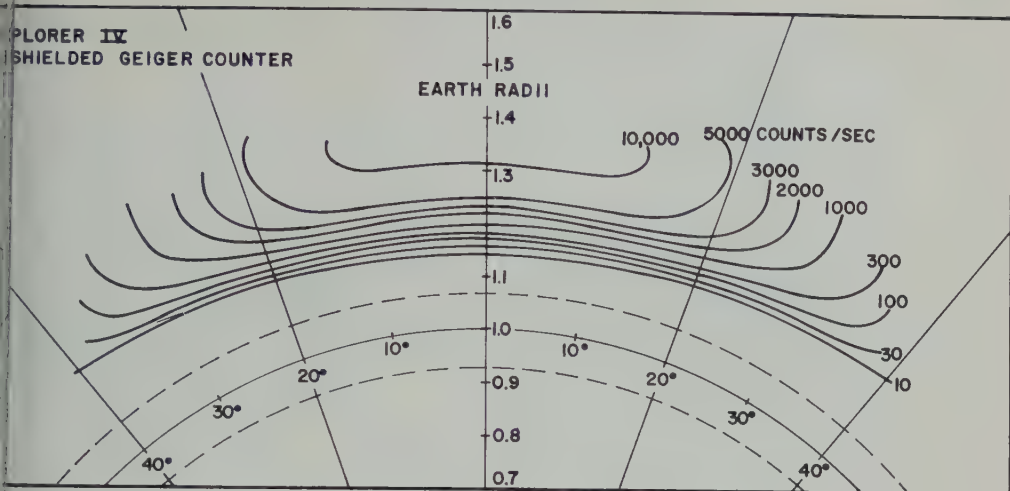


Fig. 7. The contours of constant counting rate shown in Figure 5 transformed to the polar coordinates R and λ .

obtained from recordings made at stations in North and South America, Europe, Africa, Australia, New Zealand, and Hawaii. Data obtained during the periods immediately following high altitude nuclear detonations have been added.

The general physical configuration of the inner part of the inner zone is illustrated in Figures 7 and 8 where the B versus L curves in Figures 5 and 6 have been transformed into polar coordinates in the manner previously de-

scribed. The dashed lines approximately represent the maximum excursions of the earth's surface in these coordinates.

Some of the aspects of these graphs which will be subjects for future papers include: (a) determination of angular distributions in the manner outlined by Ray [1960]; (b) the spatial dependence of proton injection and loss mechanisms; (c) the spatial dependence of the proton energy spectrum; (d) the time dependence of the intensities; and (e) comparison with

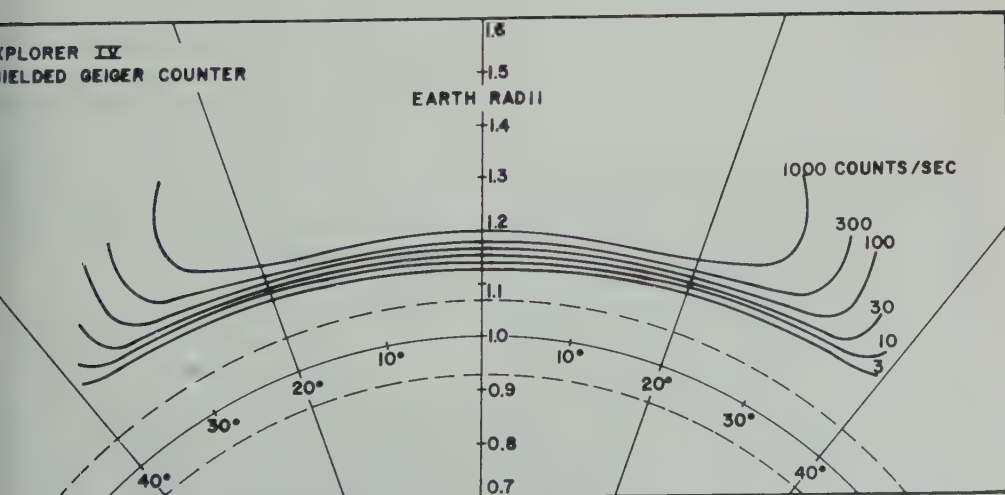


Fig. 8. The contours of constant counting rate shown in Figure 6 transformed to the polar coordinates R and λ .

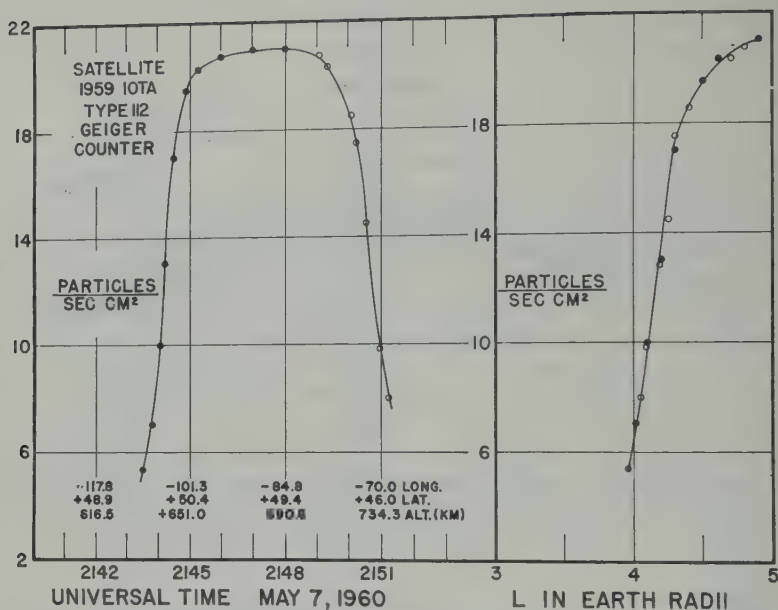


Fig. 9. The left side of this figure shows the omnidirectional intensity measured by the shielded Geiger counter aboard Explorer VII as a function of time during a solar cosmic ray event. The right side of the figure shows the same data points plotted versus L . On some occasions the modification of the earth's field by external current systems permits solar particles to arrive along lines of force labeled by considerably lower L values.

data obtained by instruments on other space vehicles.

Solar cosmic rays. The application of the B and L coordinate system is by no means limited to trapped particles; it can be usefully employed in the study of any phenomenon in which motion tends to be along lines of force. Some obvious examples are auroral particles, whistler mode wave propagation, and low energy cosmic rays.

The latitude dependence of solar cosmic rays is of particular interest in that it can be used to obtain information about both the energy spectrum of the particles and the distortion of the earth's magnetic field by external current systems. An example of the latitude dependence of solar cosmic rays is shown in Figure 9. The left side of this figure shows the intensity versus time measured by the shielded Geiger counter aboard Satellite 1959 (Explorer VII) during a solar cosmic ray event, and the right side shows the same data points plotted versus L .

Time-varying magnetic field. The magnetic shells labeled by L values greater than 3 earth radii are probably distorted by an important

amount during magnetic storms. It would be difficult if not impossible to calculate true instantaneous values of I during a magnetic storm. It is suggested that the values calculated for magnetically quiet periods be used as a reference system in which the perturbations of the various phenomena can be studied.

Computation of I . Direct evaluation of $I = \oint (1 - B_i/B)^{1/2} ds$ entails a considerable amount of computation. A number of relatively fast methods of computing I have been devised. At least one of these methods (which in general involve various means of interpolation) will probably be made available in the form of a computer program written in the Fortran language.

Conclusion. The choice of B and L as primary coordinates is, of course, quite arbitrary. One possible alternate for B is B/B_0 or $L^3 B/M$ where $B_0 = M/L^3$ is the equatorial value of B on the line of force. It should be noted that any pair of the quantities B , L , R , λ , and I can be used in place of B and L . To avoid confusion and to facilitate the comparison of different sets of data, it is suggested that a pa-

tes other than B and L be employed
a substantially improved presentation is

acknowledgments. I would like to thank Pro-
f. A. Van Allen and Professor E. C. Ray
for suggestions and helpful discussions. I
also like to thank Captain J. A. Welch,
D. C. Jensen, and others at the Air Force
Weapons Center for furnishing many of
the necessary computations.

This research is supported in part by the Office
of Naval Research (contract number Nonr 93803)
and the National Aeronautics and Space Ad-
ministration (contract number NASw-17).

REFERENCES

D. C., and W. A. Whitaker, A spherical
harmonic analysis of the geomagnetic field,

J. Geophys. Research, **65**, 2500, 1960.

Northrup, J. G., and E. Teller, Stability of the
adiabatic motion of charged particles in the
earth's magnetic field, *Phys. Rev.*, **117**, 215-225,
1960.

Pennington, R. H., Equation of a charged particle
shell in a perturbed dipole field, *J. Geophys.*
Research, **66**, 709-712, 1961.

Ray, E. C., On the theory of protons trapped in
the earth's magnetic field, *J. Geophys. Research*,
65, 1125-1134, 1960.

Van Allen, J. A., C. E. McIlwain, and G. H. Lud-
wig, Radiation observations with satellite 1958e,
J. Geophys. Research, **64**, 271-286, 1959a.

Van Allen, J. A., C. E. McIlwain, and G. H. Lud-
wig, Satellite observations of electrons artifi-
cially injected into the geomagnetic field, *J.*
Geophys. Research, **64**, 877-891, 1959b.

(Manuscript received August 28, 1961.)

A Theoretical Model of Temperature Variations at the Surface of an Orbiting Satellite

HEINZ H. LETTAU

*Department of Meteorology
University of Wisconsin, Madison*

Abstract. A closed solution of an elementary thermal diffusion equation for solid semi-infinite conductors is presented which satisfies the black-body-radiation boundary condition of proportionality between boundary heat conduction and the fourth power of surface temperature, as imposed by Stefan's law. Theoretical heating and cooling trends are shown to agree satisfactorily with empirical data measured with the aid of Suomi's bolometric equipment onboard Explorer VII for passages into and out of the earth's shadow.

Introduction. Variations in skin temperature of an artificial satellite are primarily caused by passages into and out of the earth's shadow. The resulting cooling and heating trends are controlled by the orbit elements, the intensities of solar and terrestrial radiation fluxes, and a combination of several parameters of the satellite material and its surface structure, including thermal conductivity, emissivity, and heat capacity. The theoretical problem, basically, is to solve the equation of thermal diffusion in a variable state based on the assumption that the heat flux at the surface of a solid conductor is proportional to the fourth power of surface temperature. This boundary condition is imposed by Stefan's law.

As stated in the literature that an exact closed solution of the heat diffusion problem for a semi-infinite solid conductor with the boundary condition has not been found [Carslaw and Jaeger, 1947, p. 15]. According to a theoretical prediction of temperature variations of an artificial satellite has been obtained by numerical integration of the differential equation of thermal diffusion with the aid of an electronic computer for assumed solar and terrestrial radiation fluxes, orbit elements, and material parameters; reference can be made to Heller and Jones [1960]. A simpler and more direct method of predicting satellite temperature trends could be highly useful, especially for the bolometer problem. This is the reversed problem, that is, the determination of solar and terrestrial radiation fluxes from recorded heating and cooling rates, for known satellite alti-

tudes and attitudes, and material parameters.

Lettau [1951] has discussed an approximate solution of the basic problem, in connection with surface temperature variations on the moon during a lunar eclipse. During recent years, highly reliable and detailed measurements of cooling and heating trends of artificial satellites have been obtained. Especially noteworthy are data for Explorer VII; reference can be made to Heller and Jones [1960] and Suomi [1960]. It will be demonstrated (see Fig. 1) that satisfactory agreement exists between Suomi's records and theoretical trends obtained from a closed integration of a simple differential equation of the basic problem. A brief review of the author's theoretical model appears to be in order.

Theoretical background. Inside a semi-infinite solid conductor the independent variables of the problem are time t (sec) and distance z (cm) normal to the surface. The dependent variables are temperature T (deg Kelvin), and heat flux F (ly/sec) in the direction normal to the surface. Let the dot indicate partial differentiation with respect to time t , and a prime partial differentiation with respect to depth z . In the absence of internal heat sources or sinks the basic equation of heat continuity is

$$F' = -C\dot{T} \quad (1)$$

where C = volumetric heat capacity (ly cm⁻¹ deg⁻¹). The second basic equation is Fourier's law,

$$F = -\lambda T' \quad (2)$$

where λ = heat conductivity (ly cm sec⁻¹ deg⁻¹).

Let the discussion be restricted to the case of a homogeneous conductor, that is, for $\lambda' = C' = \dot{\lambda} = \dot{C} = 0$. Then, the thermal diffusivity $\kappa = \lambda/C$ (cm²/sec) is a constant, and (1) and (2) yield the two elementary diffusion equations of the problem,

$$\dot{F} = \kappa F'' \quad (3a)$$

$$\dot{T} = \kappa T'' \quad (3b)$$

For any arbitrary frequency n (sec⁻¹) a particular solution is

$$F = b_{n,0} \exp [int - z\sqrt{in/\kappa}] \quad (4)$$

where $b_{n,0}$ is an amplitude value at the surface ($z = 0$), and $i = \sqrt{-1}$. It follows from differentiations of (4) that

$$F' = -F\sqrt{in/\kappa} \quad (5a)$$

$$\dot{F} = inF \quad (5b)$$

Solving equation (1) for \dot{T} and combining it with (5a) produces

$$\dot{T} = -F'/C = F\sqrt{in}/(C\sqrt{\kappa}) \quad (6a)$$

$$\dot{T} = F\sqrt{in/\lambda C} \quad (6a)$$

$$\dot{T}^2 = F^2 in/\lambda C \quad (6b)$$

In the squared form (equation 6b), the frequency n can be eliminated with (5b), whereupon

$$(\dot{T})^2 = F\dot{F}/\lambda C \quad (7)$$

For surface conditions the parameter $\sqrt{\lambda C}$ is the 'thermal admittance' (ly deg⁻¹ sec^{-1/2}) of the conductor. The reciprocal value $(\lambda C)^{-1/2}$ is sometimes referred to as the 'thermal inertia' constant [Bobrovnikoff, 1959]. In the following section (7) will be considered at $z = 0$; at this level, let $T(z, t) = T_0(t)$, and $F(z, t) = F_0(t)$.

For brevity the above derivation has been restricted to the simple case of a harmonic particular solution in F . However, this is a sufficient but not a necessary condition. It can be seen that (7) is valid whenever $F/F' = F'/F''$. The last requirement is satisfied for harmonic and other solutions. In the original discussion [Lettau, 1951] the general integral of the thermal diffusion equation was considered, and it was found that (7), at the level $z = 0$, produces a tolerable approximation to actual surface temperature variations of any time scale.

Surface heat-budget equation. In the absence of an atmospheric envelop, when no additional heat-consuming processes are present (such as surface melting or freezing, condensation, evaporation, photochemical transformations), the heat budget at the surface of a semi-infinite conductor is described by the equation

$$F_0 = (1 - \alpha)I_0 - \epsilon\sigma T_0^4$$

where α = surface albedo, I_0 = normal component of incoming radiation flux density (ly/sec), ϵ = surface emissivity, and σ = Stefan constant = $1.38 \cdot 10^{-12}$ (ly sec⁻¹ deg⁻⁴). Let us define a radiation equilibrium temperature, T_* , by the following identity:

$$\epsilon\sigma T_*^4 \equiv (1 - \alpha)I_0$$

With the aid of T_* equation 8 can be reformulated as

$$F_0 = \epsilon\sigma(T_*^4 - T_0^4)$$

For the simplified case of $\dot{I}_0 = \dot{T}_* = 0$ it follows from a differentiation of (10) that

$$\dot{F}_0 = -4\epsilon\sigma T_0^3 \dot{T}_0$$

For convenience, let $a^2 \equiv 4(\epsilon\sigma)/\lambda C$. Writing (7) for surface conditions, and substituting expressions (10) and (11) for F_0 and \dot{F}_0 produces

$$\dot{T}_0^2 = -a^2(T_*^4 - T_0^4)T_0^3 \dot{T}_0 \quad (11a)$$

$$\dot{T}_0 = -a^2(T_*^4 - T_0^4)T_0^3 \quad (11b)$$

T_0 in (12a) can be replaced by iT_0 , whereupon

$$(i\dot{T}_0)^2 = -a^2[T_*^4 - (iT_0)^4](iT_0)^3(i\dot{T}_0) \quad (12a)$$

or

$$\dot{T}_0 = a^2(T_*^4 - T_0^4)T_0^3 \quad (12b)$$

considering that $i^2 = -1$ and $i^4 = 1$.

It can readily be verified that only (12b) produces physically realistic heating and cooling trends. That is, cooling conditions ($\dot{T}_0 < 0$) are given when for a certain period ($0 \leq t \leq t_h$) the radiation equilibrium temperature (T_*) is lower than the initial surface temperature ($T_{00,c}$, that is, T_0 at time $t = 0$). Correspondingly, a heating period ($0 \leq t \leq t_h$) is characterized by $\dot{T}_0 > 0$, for $T_{*,h} > T_{00,h}$. Equation 13b produces $\dot{T}_0 < 0$ for $T_{*,c} < T_{00,c}$, and also $\dot{T}_0 > 0$ for $T_{*,h} > T_{00,h}$, whereas (12a) would result in reversed signs of \dot{T}_0 . Therefore, in the following section, the discussion of skin temperature variations

will be based on (13b) for both cooling and heating trends.

Ideal satellite temperature trends. For simplicity, an abrupt start of an individual cooling or heating period will be assumed, beginning with the instant of every passage into or out of the earth's shadow. In a first-order approximation the radiation equilibrium temperatures ($T_{*,c}$ or $T_{*,h}$, respectively) will be considered constant for a given cooling or heating trend. In other words, certain second-order effects caused by actual temporal and spatial variations in both terrestrial and solar radiation as well as satellite altitude and attitude, are neglected.

Equation 13b is an ordinary differential equation for the surface temperature as a function of time. If T_* at $t > 0$ is constant, a closed integration can be performed. For brevity, let us use dimensionless dependent and independent variables, denoted by x and y in the following relations:

$$x = (T_0/T_{00})^2 \quad \text{and} \quad y = t/\tau \quad (14)$$

where reference temperature T_{00} is the initial value of the surface temperature (i.e., the temperature at or just before the time of passage), and τ = reference time interval which depends on the initial or reference temperature, Stefan's constant, and the thermal capacity of the satellite material in the form

$$\tau = \lambda C / (\epsilon \sigma)^2 T_{00}^6 = 4/a^2 T_{00}^6 \quad (15)$$

we also define a dimensionless parameter p , which is a characteristic of the total intensity of cooling or heating trends:

$$p = (T_*/T_{00})^2 \quad (16)$$

In the aid of the above abbreviations, and requiring that $d(1/x) = -dx/x^2$, equation 13b can be reformulated as follows:

$$dx/x^2(p^2 - x^2) = 8 dy \quad (17)$$

Equation 17 is solved for cooling trends (that is, $p < 1$) by

$$1 - x^{-1} + (2p_c)^{-1} [\log_e (x + p_c)(1 - p_c) - \log_e (x - p_c)(1 + p_c)] = 8p_c^2 y \quad (18a)$$

for heating trends (that is, for $p = p_h > 1$) by

$$1 - x^{-1} + (2p_h)^{-1} [\log_e (p_h + x)(p_h - 1) - \log_e (p_h - x)(p_h + 1)] = 8p_h^2 y \quad (18b)$$

For the numerical verification of the solution it is most practical to start from an assumed series of x values ranging from 1 to p and to compute y for each of these x values. Then y is converted to time t , and x to skin-temperature T_0 , with the aid of the defining equations (14).

Lettau [1951] has used forms corresponding to equations 18 for the study of lunar surface temperature variations, with step functions for T_* in view of the rather gradual change of radiation at the center of the moon's disk during a lunar eclipse. Note that for the ideal, or extreme, condition of cooling in the absence of any incoming radiation (that is, for the special case of $p_c = T_{*,c} = 0$, at $t > 0$) equations 17 and 18 reduce to

$$-x^{-4} dx = 8 dy \quad \text{or} \quad x^{-3} = 1 + 24y \quad (19)$$

or

$$T_0 = T_{00}(1 + 24t/\tau)^{-1/6}$$

The skin temperature varies inversely proportional to the sixth root of time. This specific temporal function can be considered the main characteristic of the heat diffusion process with radiation boundary condition. A related fact is that the reference time interval τ (which determines the relative time scale of heating or cooling trends for finite T_* values) is inversely proportional to the sixth power of the initial temperature T_{00} (see equation 15). Consequently, for a given absolute value of the difference between initial temperature and equilibrium temperature, the approach to radiation equilibrium is much more rapid for the heating trend than for the cooling trend.

Comparison with Explorer VII data. The temperature sensing equipment on board Explorer VII includes a set of bolometers in the form of thin hemispheric shells made of silver, supported at the inside near the pole by a slender mounting post. Thermistors are employed to record the temperature near the point where the mounting post joins the silver shell. For detail of the instrumentation and data recording, reference is made to Suomi [1960]. In addition to tentative data tabulations in the 1960 report, V. E. Suomi has kindly supplied the author with two time series of temperature measured by the 'white sensor,' for constant sampling intervals

TABLE 1. Examples of Heating and Cooling Trends Recorded by the 'White Sensor' On Board Explorer VII

t	T_0	\dot{T}_0	$\frac{\dot{T}_0}{T_0^3}$	Δ	T_0^4	Δ'	(a^2)	$\frac{\dot{T}_0}{T_0^3 a_m^2}$	T_*^4	$T_0^4 - T_{*m}^4$	a^2
00	204.8
42	205.2
84	219.7	.368	350	...	233	291	524	-271	1.2
126	236.6	.334	250	-208	313	142	1.46	208	521	-191	1.3
168	247.4	.215	142	-164	375	107	1.53	118	493	-129	1.1
210	254.5	.143	86	-89	420	77	1.15	71	491	-84	1.0
252	259.3	.094	53	-54	452	53	1.02	44	496	-52	1.0
294	262.3	.058	32	-34	473	34	1.00	27	500	-31	1.0
336	264.4	.036	19	...	486	16	502	-18	1.0
378	265.3
										504	...
00	255.4
42	239.6	-.315	-229	...	329	191	138	199	1.1
84	228.9	-.198	-165	113	275	-82	1.38	137	138	145	1.1
126	223.0	-.128	-116	60	247	-49	1.22	97	150	117	0.9
168	218.1	-.110	-105	21	226	-38	0.58	88	138	96	1.0
210	213.8	-.093	-94	21	209	-29	0.73	78	131	79	1.1
252	210.3	-.078	-84	20	197	-25	0.80	69	128	67	1.2
294	207.2	-.066	-74	19	184	-21	0.90	62	122	54	1.3
336	204.8	-.056	-65	19	176	-16	1.19	54	122	46	1.4
378	202.5	-.045	-55	...	168	46	122	38	1.4
420	201.0
										130	...

Notes:

- t = time in sec.
- T_0 = sensor temperature in deg Kelvin.
- \dot{T}_0 = time rate of temperature change in deg/sec.
- \dot{T}_0/T_0^3 and overlapping differences Δ in $10^{-10} \text{ sec}^{-1} \text{ deg}^{-2}$.
- T_0^4 and overlapping differences Δ' in 10^7 deg^4 .
- (a^2) = tentative values of the coefficient in equation 21, in $10^{-17} \text{ deg}^{-6} \text{ sec}^{-1}$.
- a_m^2 = assumed average of all (a^2) values = $1.2 \cdot 10^{-17} \text{ sec}^{-1} \text{ deg}^{-6}$.
- $\dot{T}_0/T_0^3 a_m^2$, T_*^4 , and $T_0^4 - T_{*m}^4$ in 10^7 deg^4 , where T_{*m} = average radiation equilibrium temperature estimated separately for heating and cooling trends.
- a^2 = value of the coefficient in equation 20.

of 42 seconds. This information is listed in Table 1 and illustrated in Figure 1. It consists of an example of a typical warming trend (that is to say, approximately 7 minutes of record during a passage out of the earth's shadow, on October 29, 1959), and also an example of a typical cooling trend (approximately 10 minutes). The cooling trend is a representative composite derived by the University of Wisconsin group for several passages of Explorer VII into the earth's shadow, for minimum conditions of radio noise and other disturbances.

Figure 1 shows that in general the empirical data agree quite satisfactorily with theoretical

trends based on equations 18, for the indicated values of initial temperature and radiation equilibrium temperature. Note the characteristic difference in time rates of temperature change for heating and cooling. This fact is explained in the preceding section. Figure 1 shows that the heating trend reaches equilibrium, practically within less than 5 minutes, while the cooling trend shows that even after 30 minutes the temperature can be still approximately 1° above the radiation equilibrium. In detail, considerable second-order discrepancies between empirical and theoretical trends can exist.

For more specific testings of the theory,

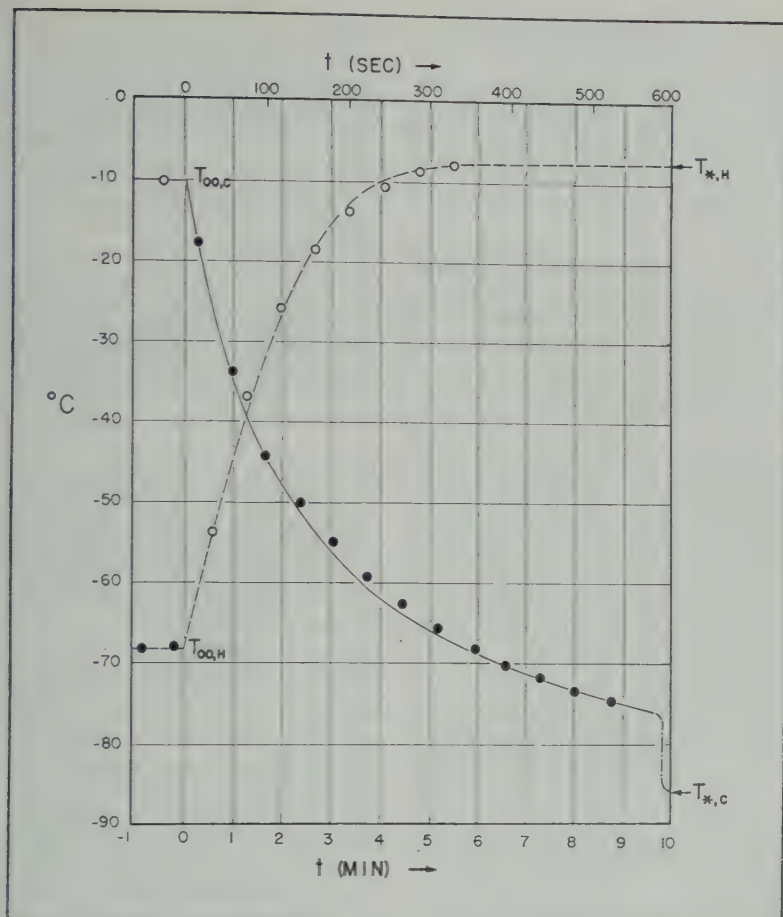


Fig. 1. Examples of cooling and heating trends recorded by the 'white sensor' on board Explorer VII.

ed for the coefficient a^2 , which is the parameter of the sensor:

$$a^2 = -(\dot{T}_0/T_0^3)/(T_0^4 - T_*^4) \quad (20)$$

ation can only be used for the computation of a^2 when T_* is known. However, since a^2 has been derived for the special case of $\dot{T}_* = 0$, it is also that

$$(a^2) = -\Delta(\dot{T}_0/T_0^3)/\Delta T_0^4 \quad (21)$$

where the symbol Δ indicates finite differences with respect to time. Equation 21 can be directly applied to observational data. Results of two computations are listed in Table 1. It is seen that (21) produces a^2 values which show considerable scatter about an average of approximately $1.2 \cdot 10^{-17} \text{ sec}^{-1} \text{ deg}^{-4}$. The large scatter is obviously due to the observational

error of the second-order differences in (21). However, the mean value $(a^2)_m$ may be used to estimate average T_* values, employing the following rewritten form of (13b):

$$T_*^4 = T_0^4 + (\dot{T}_0/T_0^3)/(a^2)_m \quad (22)$$

The two resulting $(T_*)_m$ values for the heating and the cooling trend, respectively, are indicated in Table 1, also a subsequent recomputation of a^2 with the aid of $(T_*)_m$ and equation 20. In comparison with the tentative values for (a^2) based on (21) the last column shows a satisfactorily reduced scatter. A systematic increase of a^2 for decreasing sensor temperature is also indicated.

Truly significant results can be expected only on the basis of an extensive investigation of a large amount of satellite data. This work is

presently in progress at the Department of Meteorology of the University of Wisconsin under the direction of V. E. Suomi. Its main objective is the study of terrestrial radiation fluxes, that is, to obtain T_* data along the orbit paths in the earth's shadow as a function of geographic latitude and longitude, season, or weather pattern.

The tentative value of a^2 assumed in Table 1 corresponds to $a = 0.346 \cdot 10^{-8} \text{ sec}^{-1/2} \text{ deg}^{-1}$. When converted to thermal admittance, assuming that $\epsilon = 1$, it follows for the white sensor:

$$\sqrt{\lambda C} = 2\sigma/a = 0.000,80 \text{ ly deg}^{-1} \text{ sec}^{-1/2}$$

This relatively low value of the effective thermal admittance is explained by the construction of the bolometer, especially the thin hemispheric shell supported by a slender mounting post. Obviously, a low value of thermal admittance is favorable for bolometric measurements.

Conclusions. It must be emphasized that the elementary differential equation of the problem (equation 7) was derived for the special case of the homogeneous, semi-infinite conductor. These assumptions are hardly satisfied by structures of relative complexity such as Suomi's satellite bolometers; moreover, the relatively large temperature range may result in significant changes of volumetric heat capacity and/or heat conductivity of the material. In view of these facts, the agreement between theoretical and empirical results, as evidenced by the curves in Figure 1 and the data in Table 1, is surprisingly close. This can be explained by the use of an effective

value for the thermal admittance, derived from the actual response of bolometer temperature radiation changes. The method may be referred to as in-flight calibration. It can be expected that the elementary model of heat diffusion with black-body-radiation boundary conditions will prove of value in the general problems of bolometry.

Acknowledgments. The research reported in this paper was supported by the U. S. Army Electronic Proving Grounds, Meteorology Department, Fort Huachuca, Arizona, under contract DA-039-SC-80282. I appreciate the stimulating and helpful suggestions extended by V. E. Suomi and F. B. House.

REFERENCES

- Bobrovnikoff, N. T., Natural environment of the moon, *WADC Phase Tech. Note 3, Contract 33(616)-5914*, Ohio State University Research Foundation, Columbus 8, Ohio, 1959.
- Carslaw, H. S., and J. C. Jaeger, *Conduction of Heat in Solids* 1st ed., Oxford University Press, 1947.
- Heller, G., and B. P. Jones, Thermal properties of Explorer VII, chap. 6 in *Juno II Summary Project Report*, vol. 3, *Natl. Aeronaut. and Space Admin., Rept. MTP-M-RP-60-1*, Huntsville, Ala., 1960.
- Lettau, H., On the heat budget of the moon and the surface temperature variation during lunar eclipse, *Geofis. pura e appl.*, 19, 1, 1959.
- Suomi, V. E., The thermal radiation balance experiment on board Explorer VII, chap. 11 in *Juno II Summary Project Report*, vol. 3, *Natl. Aeronaut. and Space Admin., Rept. MTP-M-RP-60-1*, Huntsville, Ala., 1960.

(Manuscript received July 9, 1961;
revised August 29, 1961.)

Propagation Characteristics of Whistlers Trapped in Field-Aligned Columns of Enhanced Ionization

R. L. SMITH

*Radioscience Laboratory, Stanford University
Stanford, California*

Abstract. Evidence from whistlers shows that the outer ionosphere contains columns or ducts of enhanced ionization. The theory of propagation in these ducts shows an upper cutoff frequency at one-half the gyrofrequency. The average propagation velocity for whistlers trapped in the ducts can be approximated by assuming that the energy follows along the ionization maximum with wave normals aligned with the magnetic field.

INTRODUCTION

It has been shown that whistler energy may be trapped by field-aligned columns or ducts of enhanced ionization in the outer ionosphere [Helliwell, and Yabroff, 1960]. In this paper we shall discuss the evidence from whistlers which supports the hypothesis that such ducts exist. The theory of propagation in the ducts will be expanded to show the time-delay effects and an upper cutoff frequency for whistler propagation. The results show that, for frequencies below half the minimum gyrofrequency, the propagation velocity may be approximated by assuming that the energy travels along the ionization maximum with wave normals aligned with the magnetic field. This result greatly simplifies the analysis of whistlers [Helliwell, Crary, Pope, and Carpenter, 1956]. A cutoff frequency of approximately one-half the minimum gyrofrequency is predicted for whistlers propagating in the ducts.

EVIDENCE FOR DUCTS

Data from whistler observations that indicate the presence of columns of enhanced ionization are listed below. The succeeding sections show that these data cannot be explained if the outer ionosphere is smoothly varying and free of irregularities capable of supporting trapped propagation paths. We will show, however, that the principal features of whistlers can be explained by assuming the existence of field-aligned ionization.

Trapped gliding tones. Some whistlers sound like gliding tones. The spectrograms of

such whistlers often reveal very thin traces. The time delay of a whistler from its originating atmospheric is approximately 1 second. Spectrograms may indicate trace widths of the order of 0.01 second at a given frequency.

On spectrographic analysis most whistlers show a number of distinct components whose traces are often very thin. The components usually have a common origin. The collection of components may then be designated a 'whistler group.' In a nose whistler group the nose frequency or frequency of minimum time delay usually decreases systematically with time. Nose whistler components usually exhibit sharply defined traces. Such behavior would not be expected on the basis of the theory of propagation of whistlers in a homogeneous medium [Smith, 1960a].

2. Repetition of components. Different whistlers in the same recording period (2 minutes every hour is the usual sampling schedule) almost always have components that show exactly the same time delay and are similar in general appearance. These components may have varying ratios of intensity from one whistler to the next, but a strong whistler will almost invariably contain all the components of a weaker whistler in the same period. The same components often appear in succeeding hours [Carpenter, 1960]. The data thus suggest many isolated and persistent paths of propagation.

3. Low decrement of echoes. Whistler echo trains with very little decrease in amplitude from one echo to the next have been observed on occasion. The whistler is apparently confined so that the energy does not diverge in space.

Such a confinement might result from ducts of enhanced ionization.

4. *Multiple echo delays.* The time-delays of certain components of whistlers are, within experimental error, usually related to those of their echoes by exact even or odd integral multiples. The experimental error is often less than 1 per cent. Furthermore, in a given 2-minute period of observation the ratio of time delays of components of short whistlers to those of long whistlers is nearly always 1:2. A component that is found to echo is, however, not necessarily the strongest component of the one-hop, or short, whistler.

5. *Echo coupling.* The only exception to the above rule of multiple echo delays is that coupled echoes, which appear to be combinations of two or more components of the one-hop whistler, are sometimes observed. For example, a short whistler may contain two strong components with time delays of A and B at a given frequency. The three-hop group may contain not only components at $3A$ and $3B$ but also at $2A + B$ and $2B + A$. Examples of such echo combinations were first published by *Morgan, Curtis, and Johnson* [1959]. Other examples have been noted on records from the IGY Whistlers-West program [*Helliwell and Carpenter*, 1961]. Echo coupling strongly suggests the existence of isolated and unique paths of propagation in the outer ionosphere.

6. *Nose whistler upper cutoff frequencies.* For all nose whistlers that have been analyzed, the ratio of the upper cutoff frequency to the estimated minimum gyrofrequency has never exceeded 0.56 [*Smith*, 1960b].

7. *Similar whistlers from different source locations.* *Helliwell, Taylor, and Jean* [1958] found the location of whistler-causing atmospheres by direction-finding and triangulation. The results showed that whistlers at Stanford and Boulder with similar component structure were caused by lightning from the Pacific Ocean, the Gulf of Mexico, and North Dakota. The dispersion properties of particular components does not depend on the location of the source.

A similar result can be deduced from the 'hybrid' whistler [*Helliwell*, 1959], which consists of a short whistler and an associated long whistler, both excited by the same lightning stroke. The lightning source occurs in the same hemisphere as the observing station. When the

proper propagation delays are taken into account, the time delays of certain of the components of the long whistler are related by exactly 2:1 to the components of the short whistler.

8. *Whistler occurrence rate as a function of geomagnetic latitude.* The rate of whistler occurrence is known to increase with increasing geomagnetic latitude from the equator to at least 50° [*Helliwell and Carpenter*, 1961; *Crouchley*, 1961]. The occurrence rate is highly variable at higher latitudes and depends in part on the particular observing site.

9. *Absence of leading components in echoes.* It has been noticed in the study of whistler echoes that the expected echoes of the leading components of short whistlers are sometimes absent in the echo structure, although echoes of weaker components showing greater time delays are often present.

SMOOTH OUTER IONOSPHERE MODELS

To explain the above observations we will first consider the possibility that the density in the outer ionosphere decreases smoothly with distance, with no irregularities capable of trapping whistler energy. A number of studies have been made using ray-tracing techniques to determine the paths of whistlers for various initial conditions and different smooth models of the outer ionosphere. Using the results of these studies we let us consider various factors that might account for the fine structure noticed in whistler spectrograms.

Yabroff [1959] has shown for a certain model that over a wide range of initial latitudes the resulting low-frequency whistler tends to focus at one latitude. Furthermore, the energy tends to arrive at the same time for the same variation in initial latitude. However, this type of behavior could not explain multiple components.

One suggested explanation for multiple-component whistlers is that the excitation conditions vary with distance from the source. The intensity of the signal from a vertical radiator is zero directly overhead. A maximum signal intensity would be expected at the lower ionosphere in the neighborhood of 100 km from the source. It is expected that multiple reflections between earth and ionosphere will cause the total whistler-mode signal to fluctuate with distance from the source. However, variation in excitation

cannot account for the repetition of whistlers that have sources in different places. The existence of hybrid whistlers clearly shows that the variation in source location is not the cause of discrete whistler components.

Helliwell, Crary, Pope, and Smith [1956] suggested, as another possibility, that the fine structure might be related to irregularities in the lower ionosphere that focus whistler energy. The irregularities may be found to exist near the receiving end or the starting end of the main whistler path. The existence of a structure of similar effect may be suggested from a recent paper by Budden [1960]. He suggests that whistler energy will penetrate the ionosphere in the presence of irregularities of ionization that have dimensions very small when compared with a wavelength unless the irregularities are elongated in the direction of the magnetic field. We may assume that these elongated irregularities are found in localized regions, thus producing ducts in the lower ionosphere through which whistler energy may pass.

If we are assuming for the moment that the outer ionosphere is smooth, we would expect the location of the irregularities in the ionosphere to be uncorrelated in the two hemispheres. Then the components of long and short whistlers would be uncorrelated. This is in contradiction to the observed facts.

Calculations made by Yabroff for smooth ionosphere models indicate that the ray path depends on the initial wave normal angle, and the final wave normal angle will usually be different from the initial value. For many of Yabroff's cases the final value of the wave normal at the end of the path approached 90° with respect to the magnetic field. The initial value of the wave normal for the first-hop whistler is taken to be approximately vertical. Helliwell [1953] has shown, the high refractive index of the ionosphere refracts all the waves below the ionosphere to a small cone around the vertical. The new value of initial wave normal for a second-hop whistler is found by considering the reflection of the final value of the first-hop whistler. This new value will not be close to the vertical. Thus the ray path of the echoes could not be the same as that of the first-hop whistler. Furthermore,

coupled echoes would not exhibit the simple relationship described previously.

The work of Yabroff [1959] does not indicate any sharp upper-frequency limit for whistlers when the outer ionosphere is smooth. The observed upper-cutoff frequency limit is easily explained by the theory presented below.

Smith [1960a], after studying the propagation of whistlers in a homogeneous medium, concludes that isolated nose whistler traces can be explained only if there is present some additional feature of the medium that can focus the energy and restrict the wave normal variation.

DUCTS OF ENHANCED IONIZATION

From the evidence presented above it appears unlikely that the fine structure of whistlers is caused by excitation effects or by irregularities in the lower ionosphere. Some mechanism in the outer ionosphere must, therefore, cause the discrete components in whistlers. A likely mechanism is the presence of field-aligned columns of enhanced ionization which can trap whistler energy. Smith, Helliwell, and Yabroff [1960] show that only a small increase of ionization in the column over the background level is sufficient to trap completely all whistler energy within a given range of wave normal angles. The limit of the ray path is then defined by the boundaries of the enhanced column. Furthermore, as we shall show later, the effective group velocity of the energy is almost completely independent of the wave normal angles, ray paths, and trapping conditions within the column. Each column or duct can then easily account for an isolated component of a whistler group. Each component would be observed as a thin trace on the spectrogram.

The explanation for the absence of leading components in echoes may be simply that for the latter components a more favorable trapping angle is encountered upon reflection. The latter components represent whistlers traveling at higher geomagnetic latitudes. The magnetic field lines are more nearly vertical at higher latitudes, and hence the reflected wave normal angle would be expected to be nearly equal to the incident wave normal angle that presumably was within the trapping angle. At low geomagnetic latitudes the reflected wave normal may be far different from the incident wave normal. The reflected wave normal will probably lie outside

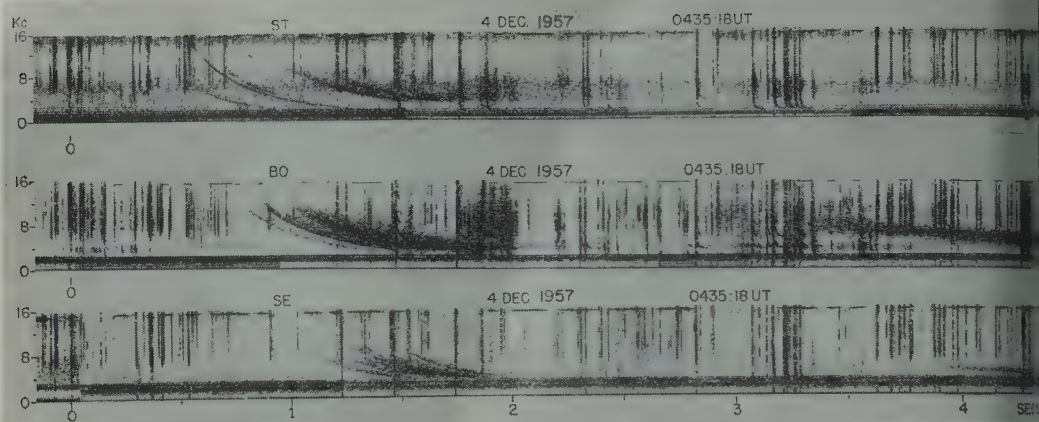


Fig. 1. Coincident whistler at Stanford, Boulder, and Seattle.

the maximum trapping angle. The above results are obtained very simply by the application of the Snell's law construction [Smith, Helliwell, and Yabroff, 1960], assuming that the ionization is horizontally stratified at the reflection point.

The delays of components in a whistler group are found to be independent of either the source location or the receiver location. The intensity of the components will, however, be a function of source and receiver location. In general, if a whistler is received simultaneously at two stations, the one at the higher geomagnetic latitude will observe relatively greater intensities in the components having greater time delays. Figure 1 demonstrates this behavior. The short whistlers shown here were received on December 4, 1957, at Stanford, California; Boulder, Colorado; and Seattle, Washington. The geomagnetic latitudes of these stations are 44° , 49° , and 54° , respectively. Examination of the atmospherics will show that the whistler was coincident at these three stations. Note that the components with greater time delay are received with more intensity at the higher geomagnetic latitudes.

The repetition of fine structure of whistlers indicates that the duct structure is relatively stable over periods of a few hours. This duct lifetime of a few hours appears to be consistent with the rate of migration of individual charged particles subjected to collisions in a magnetic field [Smith, 1960b].

Ducts with high enhancement factors should

trap all the whistler energy and thus allow propagation of whistler echo trains with little transmission loss. When the lower ionospheric layers are smooth and relatively lossless, and the lower layer is sharply bounded, most of the whistler energy should suffer total internal reflection, except for the energy having wave normals lying within a few degrees of vertical. A roughly analogous situation obtains on a lossless transmission line with very high resistance terminations at either end.

The rule of multiple echo delays follows logically from the duct hypothesis. Echo coupling probably requires that the effective duct pattern terminate sufficiently high above the reflecting level so that energy leaving one duct and reflecting from the lower layers will have sufficient space to couple to another duct. Recall that the ray direction at low frequencies must lie within a cone of $19^\circ 29'$ around the magnetic field direction.

The relation of trapping in ducts of enhanced ionization to whistler occurrence rate as a function of latitude has been discussed in a previous paper [Smith, Helliwell, and Yabroff, 1960].

VALIDITY OF RAY THEORY

In the following discussion we will assume that ray theory applies for propagation in the ducts. This requires that the dimensions of the duct be large compared to a wavelength. If the dimensions of a duct were less than a wavelength, one might expect a waveguide dispersion effect at low frequencies. Such effects have not

observed. At higher whistler frequencies, theory should be applicable simply because of the increased number of wavelengths per unit distance across the ducts.

EFFECTIVE GROUP-RAY VELOCITY OF A TRAPPED WHISTLER

The influence of the duct on the effective group-ray velocity should be examined. The normal angle of whistler energy trapped in the duct is continuously changing as the ray path makes snakelike oscillations from one side of the duct to the other. The group-ray velocity is therefore continuously changing. A given duct is known to contain a wide variety of initial wave normal angles. Furthermore, a range of electron densities is encountered by the ray on its side-to-side excursions.

If the ray path contains many side-to-side oscillations, the average group velocity along the duct will be nearly equal to the average group velocity of one ray path oscillation. We consider here only ducts of enhanced ionization. At the center of the duct the wave normal angle has its largest value, and the longitudinal component of group velocity is usually somewhat less than that of an electromagnetic wave traveling in the longitudinal direction. Near the edges of the duct the wave normal angle is close to zero, but, since the energy is now traveling in regions of lower ionization density, the group velocity is somewhat greater than that of a ray traveling down the center of the duct in the longitudinal direction. Thus the average group-ray velocity will be given approximately by the group velocity of a strictly longitudinal ray traveling down the center of the duct.

The true average group velocity can be determined only by integrating the ray-path equations and will be a function of the shape of the ionization distribution as well as the initial wave normal angle and frequency of the wave. The ray-path equations for an inhomogeneous anisotropic medium are given by *Haselgrove* [1954]. We will consider first the two-dimensional case for rays previously discussed [*Smith, Helliwell, and Coff*, 1960]. Let the constant magnetic field be directed along the x axis, and let the refractive index vary only in the y direction. The ray-path equations can be written

$$\frac{dx}{dt} = \frac{c}{\mu^2} \left(\mu \cos \theta + \frac{\partial \mu}{\partial \theta} \sin \theta \right) \quad (1)$$

$$\frac{dy}{dt} = \frac{c}{\mu^2} \left(\mu \sin \theta - \frac{\partial \mu}{\partial \theta} \cos \theta \right) \quad (2)$$

$$\frac{d\theta}{dt} = \frac{1}{\mu^2} \frac{\partial \mu}{\partial y} \cos \theta \quad (3)$$

$$\frac{dt_g}{dt} = \frac{\mu_g}{\mu} \quad (4)$$

In the above equations, dx/dt and dy/dt are the components of the ray velocity, c is the velocity of light, θ is the angle between the magnetic field and the wave normal, and t_g is the time of propagation of the energy. The phase refractive index and the group refractive index are given respectively by μ and μ_g .

The appropriate approximation for the phase refractive index for whistlers is

$$\mu = \frac{f_0}{f^{1/2} f_H^{1/2} (\cos \theta - \lambda)^{1/2}} \quad (5)$$

where

f = wave frequency

f_0 = plasma frequency = $\sqrt{80.6N}$

f_H = gyrofrequency

$\lambda = f/f_H$ = normalized frequency

θ = wave-normal angle

N = electron density (number per cubic meter)

The group refractive index is

$$\mu_g = \frac{f_0 \cos \theta}{2 f^{1/2} f_H^{1/2} (\cos \theta - \lambda)^{3/2}} \quad (6)$$

Equations 6 and 7 can be integrated immediately for all values of λ between 0 and 1, giving

$$\frac{f_0(y)}{f_0(0)} = \frac{\cos \theta_0}{(\cos \theta_0 - \lambda)^{1/2}} \frac{(\cos \theta - \lambda)^{1/2}}{\cos \theta} \quad (7)$$

where θ_0 is the wave normal angle at $y = 0$, the ionization maximum. This result could have been predicted by the use of the Snell's law construction. The trapping criteria can be obtained immediately by inserting the values of θ which gave a ray parallel to the magnetic field.

Let us consider a plasma frequency distribution across the duct given by

$$f_0(y) = f_0(0)(\cos y)^{1/2} \quad (8)$$

and by assuming that the energy travels along the ionization maximum with the wave normals aligned along the magnetic field lines. One of the explanations is that a duct of given enhancement factor inherently excludes wave normals exceeding a certain value. In addition, there is an averaging effect on the group ray velocity as previously discussed.

If the cross section of the duct is approximately circular and the ray does not cross the walls of the duct, the trapping properties are determined partly by the gradients of ionization. In the limiting case the ray path follows a helical path along the duct, and the trapping condition depends only on the gradient of ionization. The up ray is traveling in a region of ionization less dense than at the ionization maximum; thus it tends to travel faster. On the other hand, the wave normal is always directed away from the magnetic field, thus tending to make the group ray travel slower. There is again an averaging effect on the group-ray velocity.

VARIATIONS OF TRAPPING CONDITIONS ALONG A DUCT

Without knowing the exact size of the ducts and the variation of density along the ducts we nevertheless make some further qualitative deductions about the behavior of propagation along the ducts.

We have seen [Smith, Helliwell, and Yabroff, 1960] that for frequencies below half the gyrofrequency trapping could occur in both troughs and crests (minima and maxima) of ionization in a uniform magnetic field and no variations in refractive index along the field. Consider the more realistic picture of ducts extending along the earth's dipole field. The gyrofrequency near the entrance to the ducts is approximately 1 Mc/s and is very high compared with even the highest observed whistler frequencies. The initial value of normalized frequency, $\lambda = f/f_H$, must be close to zero. For this condition the minimum required enhancement factor for trapping in a trough is very much larger than that required for trapping in a crest. It is therefore likely that if the whistler energy is trapped, it will be trapped initially in a crest of ionization.

As the whistler travels up the field lines in the duct, the gyrofrequency slowly decreases and hence the normalized frequency increases.

The refractive index will vary along the duct because both gyrofrequency and electron density are changing. Furthermore, the enhancement factor may change. One may ask if the energy will remain in the duct. Figure 2, taken from a previous paper [Smith, Helliwell, and Yabroff, 1960], is a ray-path region diagram showing lines of constant enhancement factor for a uniform duct and for various values of normalized frequency and θ_0 , the wave normal direction when the ray is at the maximum of ionization. For propagation in a crest of given enhancement factor, the cone of possible trapped wave normals increases with frequency. Alternatively, if the maximum wave normal angle is held constant and the frequency is increased, the ray will be restricted closer and closer to the maximum of ionization as the frequency increases. The example given below indicates that the maximum wave normal angle will actually decrease as the gyrofrequency increases. Then it appears that, unless the duct properties change radically, the whistler energy will remain in the duct. We will see, however, that the properties will change abruptly near a normalized frequency of one-half.

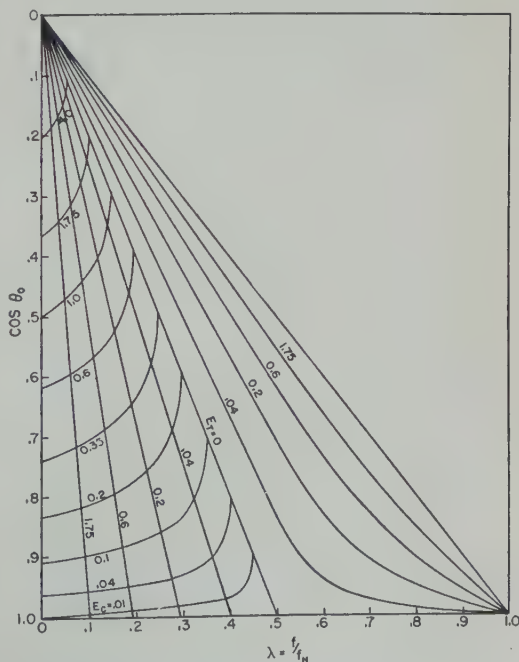


Fig. 2. Enhancement factors on the ray-path region diagram.

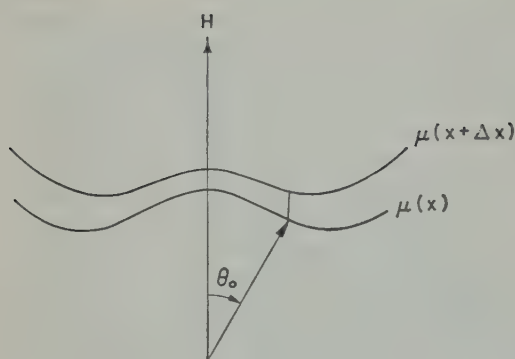


Fig. 3. Refractive index surfaces at two points along a duct.

If the whistler is trapped in a crest, the angle between the ray and the field will decrease for a given wave normal when the gyrofrequency is decreased. The maximum possible angle of the ray also decreases with decreasing gyrofrequency. If the plasma frequency decreases more rapidly than described above, it is still possible for the maximum wave normal angle to decrease with decreasing gyrofrequency. If the plasma frequency is constant, or even if it is proportional to the square root of the gyrofrequency, the refractive index will increase in the direction of decreasing gyrofrequency. This additional component of refractive index gradient tends to reduce the wave normal angle all along the ray path. The integrated effect is then a reduction in the maximum wave normal angle near the center of the column. If the plasma frequency decreases more rapidly than described above, it is still possible for the maximum wave normal angle to decrease with decreasing gyrofrequency, but this can be determined only by integration of the ray path equations.

If the ionization density does not decrease more rapidly than the gyrofrequency, the nature of the refractive index surfaces at two successive points along the duct is as indicated in Figure 3. The small change in gyrofrequency will not change the wave normal when $\theta = 0^\circ$ near the edge of the ducts. The gradient of refractive index in this case will furthermore be nearly perpendicular to the field lines. However, when the ray is at the maximum of the ionization, the gradient of refractive index will be directed entirely along the field lines, neglecting the curvature of the field lines. The wave normal angle will be close to its largest value, θ_0 , at this

point. The Snell's law construction indicates that θ_0 will decrease. Notice that if the plasma frequency were constant along the duct, it would decrease even more rapidly. The degree of enhancement necessary for trapping thus decreases with decreasing gyrofrequency. Under the actual duct enhancements decrease rapidly, the energy will remain trapped.

UPPER CUTOFF FREQUENCY

Another feature of trapping in ducts is a high-frequency cutoff near a normalized frequency of one-half. Reference to the ray-path region diagram shows that trapping can take place only in troughs of ionization for normalized frequencies above $\frac{1}{2} \cos \theta_0$. The high-frequency for which the energy may be trapped in *crests* is thus one-half the local gyrofrequency. If energy at a frequency greater than one-half the minimum gyrofrequency along the duct is traveling up the duct, then, when a critical value of gyrofrequency is reached, the energy will tend to be refracted away from the maximum of ionization. If the energy escapes from the duct, very little would be expected to return and be trapped. The observed cutoff frequency of nose whistlers is somewhat greater than one-half the minimum gyrofrequency, however. This can be explained by the following considerations. As the energy approaches the critical region, the energy will tend to follow the crest of ionization. Furthermore, the ray will be directed very close to the magnetic field. When the critical gyrofrequency has been passed, the energy will slowly refract away from the ionization maximum. If the energy does not leave the duct before it passes over the top of the path and again reaches the critical value of gyrofrequency, then the energy may remain trapped in the duct.

CONCLUSIONS

The evidence presented here shows that whistler energy is confined over most of its path in field-aligned ducts of enhanced ionization in the outer ionosphere. Furthermore, the average group velocity along the duct is given fairly accurately by the strictly longitudinal approximation; whistler energy following a snakelike path along a duct of enhanced ionization will have the same average velocity as energy directed along the field lines at the crest of ionization.

theory presented indicates that this approximation may fail at wave frequencies above the local gyrofrequency because the energy no longer be trapped in the duct. The relations describing the time delay of nose whistlers as a function of frequency are very complicated when the longitudinal group velocity can be used. These results can be used in the analysis of nose whistlers to obtain electron densities in the outer ionosphere [Smith, 1961].

Acknowledgment. This work was supported in part by the Air Force Office of Scientific Research under contract AF18(603)-126.

REFERENCES

- Allen, K. G., Effect of small irregularities on the constitutive relations for the ionosphere, *J. Research NBS*, 63D, 135-149, 1959.
- Carpenter, D. L., Identification of whistler sources from visual records and a method of routine whistler analysis, TR no. 5, contract AF18(603)-126, Radioscience Laboratory, Stanford University, AROSR-TN-60-315, March 15, 1960.
- Chapman, J., A study of whistling atmospherics, Occurrence, *Australian J. Phys.*, 14, 22-39, 1961.
- Groves, J., Ray theory and a new method of ray tracing, Report of the Physical Society Conference on the Ionosphere, Cambridge, England, 1955-364, September 1954.
- Helliwell, R. A., Hybrid whistlers (Whistler paths and electron densities in the outer ionosphere), Proceedings of the Symposium on Physical Processes in the Sun-Earth Environment, DRTE publication no. 1025, pp. 165-169, July 20-21, 1959.
- Helliwell, R. A., and D. L. Carpenter, Whistlers-West, IGY-IGC Synoptic Program, Stanford Electronics Laboratories, March 20, 1961.
- Helliwell, R. A., J. H. Crary, J. H. Pope, R. L. Smith, The 'nose' whistler—a new high latitude phenomenon, *J. Geophys. Research*, 61, 139-142, 1956.
- Helliwell, R. A., W. L. Taylor, and A. G. Jean, Some properties of lightning impulses which produce whistlers, *Proc. IRE*, 46, 1760-1762, 1958.
- Morgan, M. G., H. W. Curtis, and W. C. Johnson, Path combinations in whistler echoes, *Proc. IRE*, 47, 328-329, 1959.
- Smith, R. L., Guiding of whistlers in a homogeneous medium, *J. Research NBS*, 64D, 505-508, 1960a.
- Smith, R. L., The use of nose whistlers in the study of the outer ionosphere, TR. no. 6, contract AF18(603)-126, Radioscience Laboratories, Stanford University, no. AROSR-TN-60-861, July 11, 1960b.
- Smith, R. L., R. A. Helliwell, and I. W. Yabroff, A theory of trapping of whistlers in field-aligned columns of enhanced ionization, *J. Geophys. Research*, 64, 815-823, 1960.
- Storey, L. R. O., An investigation of whistling atmospherics, *Phil. Trans. Roy. Soc. A, London*, 246, 113-141, 1953.
- Yabroff, I. W., Computation of whistler ray paths, Stanford Research Institute Report for Stanford University, AFCRC-TN-60-71, December 1959.

(Manuscript received May 8, 1961; revised August 14, 1961.)

Properties of the Outer Ionosphere Deduced from Nose Whistlers¹

R. L. SMITH

*Radioscience Laboratory, Stanford University
Stanford, California*

Abstract. Data from a number of nose whistlers indicate that the electron density in the outer ionosphere is directly proportional to magnetic field strength. The density has an annual variation of 2:1 with a maximum near December. The average electron density is given approximately by $N = 12,000 f_H$, where N is the electron density per cubic meter and f_H is the gyrofrequency in cycles per second. Local variations as large as 40 per cent are shown to exist on occasion.

Introduction. Electromagnetic energy at very low frequencies can be propagated from one hemisphere to the other along the earth's magnetic field lines in the outer ionosphere. The scattering of free electrons disperses the impulse-like energy from a lightning discharge to form ascending gliding tones called whistlers. If the lowest frequency of a whistler is comparable to the minimum electron gyrofrequency along the propagation path, the dispersion at the lower frequencies may cause rising tones which are contiguous with the descending tones at higher frequencies. The whistler is then called a 'nose' whistler.

The advantages of whistlers in the determination of the ionization at great heights have been discussed by Storey [1958]. The time delay of a whistler from the originating discharge is proportional to the square root of ionization density, whereas other methods of determining the ionization density give quantities that are usually directly proportional to density. The conditions from regions of low ionization relative to regions of greater ionization are therefore more favorable in the whistler measurement. At low frequencies the weighting factor for the time delay of the whistler is inversely proportional to the square root of magnetic field strength. This factor increases the incremental time delay due

to ionization at 5 earth radii by a factor of roughly 11 over the time delay from an equal amount of ionization in the lower ionospheric layers. The curved shape of the ray path gives a high value of path-length per unit height near the summit of the path relative to the lower part of the path. The result of all these factors is that the electrons at the farthest distance from the earth have the greatest effect on the dispersion characteristics of whistlers.

Nose whistlers offer the additional advantage of defining the height of maximum excursion of the ray path. The frequency of minimum time delay, called the 'nose' frequency, depends primarily on the minimum gyrofrequency along the path. The nose frequency depends only to a small extent on the ionization distribution. For example, if the ionization density in the outer ionosphere did not vary with height, the nose frequency would be about 0.35 times the minimum gyrofrequency.

The nose whistler integral. Spectrographic analysis of whistlers and nose whistlers often reveals a number of pure isolated components that have a common source. Examination of whistler data and a new theory of the propagation path have led to the hypothesis that each component represents energy from the lightning source that has been trapped in a field-aligned duct of enhanced ionization in the outer ionosphere [Smith, 1961]. The data indicate that the lifetime of these ducts is a few hours. The theory suggests that enhancements of 5 per cent or greater are sufficient to explain the observed whistlers [Smith, Helliwell, and Yabroff, 1960]. The theory further indicates that the average

¹The major results of this paper were presented at the International Commission III (The Exosphere and Upper Atmosphere Region Symposium) and Commission IV of the International Scientific and Radio Union (URSI), May 1960. An abstract appears in the URSI proceedings and in Smith and Helliwell [1960].

group velocity of energy trapped in a duct can be closely approximated by assuming that the energy travels along the maximum of ionization in the duct with wave normals aligned with the magnetic field [Smith, 1961].

The theory of whistler propagation in ducts as outlined above simplifies nose whistler analysis in two ways: (1) the propagation path for a given whistler component is independent of frequency and is known to be along a field line; (2) the effective group velocity is almost entirely independent of wave normal angle so that the simplified form of group velocity for purely longitudinal waves can be used. These simplifications would not result if the outer ionosphere did not contain field-aligned irregularities [Yabroff, 1959; R. L. Smith, 1960b; Garriott, 1958].

In expressing the path of propagation of whistlers we will use the dipole approximation to the earth's magnetic field. Recent satellite results indicate that the magnetic field is very close to the predicted field out to at least 5 earth radii from the earth's center [E. J. Smith, 1960]. Nearly all whistler paths fall within this distance. The higher order terms of the magnetic field are always smaller than the dipole term and decrease more rapidly with height. The frequency-time characteristics of nose whistlers are determined principally in the region of the minimum magnetic field strength. The nose whistler can be described in terms of an effective dipole field determined from the minimum field strength at the top of the field line. This effective dipole will have an effective latitude of termination at the earth's surface. The main effect of the higher order terms is to change the termination point of the effective latitude.

The nose whistler equation can be simplified by assuming that the square of the refractive index is much larger than unity. This assumption is justified by the results. The time delay is then given by

$$T = \frac{1}{2c} \int_{\text{path}} \frac{f_0 f_H}{f^{1/2} (f_H - f)^{3/2}} ds \quad (1)$$

where

- f_0 = plasma frequency.
- f_H = gyrofrequency.
- f = wave frequency.
- c = velocity of light.

The 'dispersion' of a nose whistler is obtained by multiplying both sides of equation 1 by f and is given by

$$D = T f^{1/2} = \frac{1}{2c} \int_{\text{path}} \frac{f_0 f_H}{(f_H - f)^{3/2}} ds$$

When f tends toward zero the dispersion tends toward a constant previously defined by Storey [1953] for low frequency whistlers.

The dispersion of a nose whistler can be determined experimentally as a function of frequency. Then equation 2 can be viewed as an integral equation with the plasma frequency and the gyrofrequency (or path) as unknowns. Various theoretical methods of inverting the integral have been investigated [Storey, 1953; R. L. Smith, 1960a], but these methods have failed when applied to experimental data. The reason is simply that for the observed frequency range of whistlers the accuracy required for the simultaneous solution of the plasma frequency and the path is greater than the experimental errors.

Figure 1 shows the dispersion as a function of frequency for a well-defined nose whistler. Theoretical dispersion curves are also shown for comparison. One model in which the ionization varies as $\exp(2.5/R)$ is taken from Dunn [1954]. For the other model the ionization is directly proportional to the local gyrofrequency or magnetic field strength. The field line paths for the two models are chosen so that the resulting nose frequency is the same as that observed (14.75 kc). The paths are slightly different for the two models. The scale factors of the ionization distributions are chosen so that the dispersion at the nose frequency is the same as observed. The experimental and theoretical dispersion curves are, within experimental error, coincident. It appears then that only two significant parameters of the dispersion can be determined from a single nose whistler trace. The two parameters which will be used here are the nose frequency and the time delay from the originating discharge to the nose.

Additional information on the ionization distribution can be determined from a study of nose whistler groups. A nose whistler group consists of isolated nose whistler components excited by the same lightning flash. As discussed previously, each component presumably represents energy that has been trapped in a field-aligned

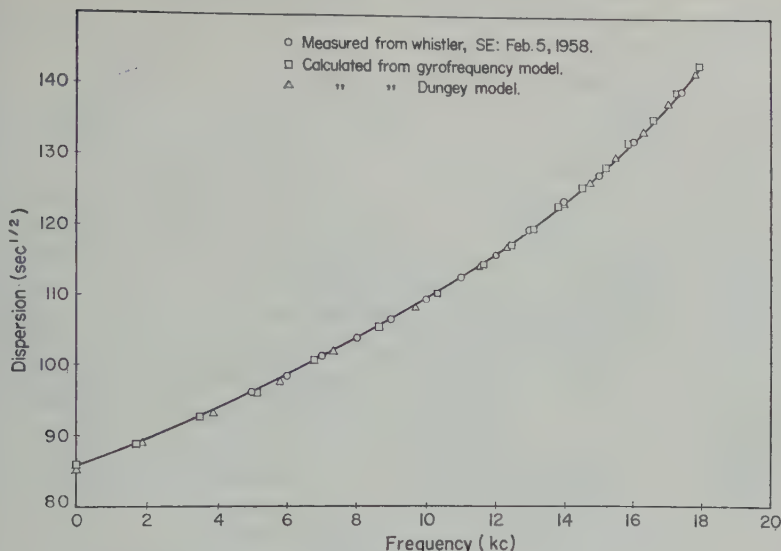


Fig. 1. Theoretical dispersion curves fitted to a nose whistler.

of enhanced ionization. Spectrographic analysis of nose whistler groups shows that components with lower nose frequencies usually arrive later than components with higher nose frequencies. The nose frequency is approximately a constant fraction of the minimum frequency along the path. Whistlers travel to regions of lower minimum gyrofrequency and travel over longer paths, and hence tend to suffer greater delays.

Nose whistler data. A number of nose whistlers covering the years 1955 to 1959 have been spectrographically analyzed. The data were obtained at stations in the northern and southern hemispheres and include a number of nose whistler groups.

The most difficult part of the scaling procedure is the determination of the time of occurrence of the whistler-causing atmospheric. This is somewhat simplified when records are available from magnetically conjugate stations. *Heath* [1960] gives several methods of identifying the sources of whistlers. After the source has been identified, the measurement of nose frequency and time delay to the nose is routine.

It is important to know whether a given whistler is short (one hop) or long (two hop). This classification can be determined readily by comparison of the time delays from the source to the whistler and its echoes, when well-defined

echoes are present. In the absence of echoes for a given whistler, one can often find similar whistlers and associated echoes received within a few hours of the desired whistler, and thus can deduce the classification. Simultaneous records from magnetically conjugate stations can also be used. The strength or type of causative atmospheric has been found to be unreliable as an indicator of short or long whistlers. Thus many short whistlers have been found with strong preceding atmospherics. The causative atmospheric may or may not show a characteristic 'tweek' waveform, depending on the distance and the earth-ionosphere wave-guide conditions.

Almost all whistlers that show noses have been found to be short. The only long nose whistler found in the present study was obtained on May 20, 1955, at Washington, D. C. (The measured time delay was therefore divided in half for comparison with the rest of the data.)

The data are shown in Figure 2 on a log-log plot. The abbreviations on the figure indicate the station where the whistlers were received: *BO*, Boulder, Colorado; *BY*, Byrd Station, Antarctica; *CO*, College, Alaska; *DU*, Dunedin, New Zealand; *SE*, Seattle, Washington; *ST*, Stanford, California; *UN*, Unalakleet, Alaska; *WA*, Washington, D. C.; *WE*, Wellington, New Zealand. It can be deduced from the data that the location of the receiving station bears no

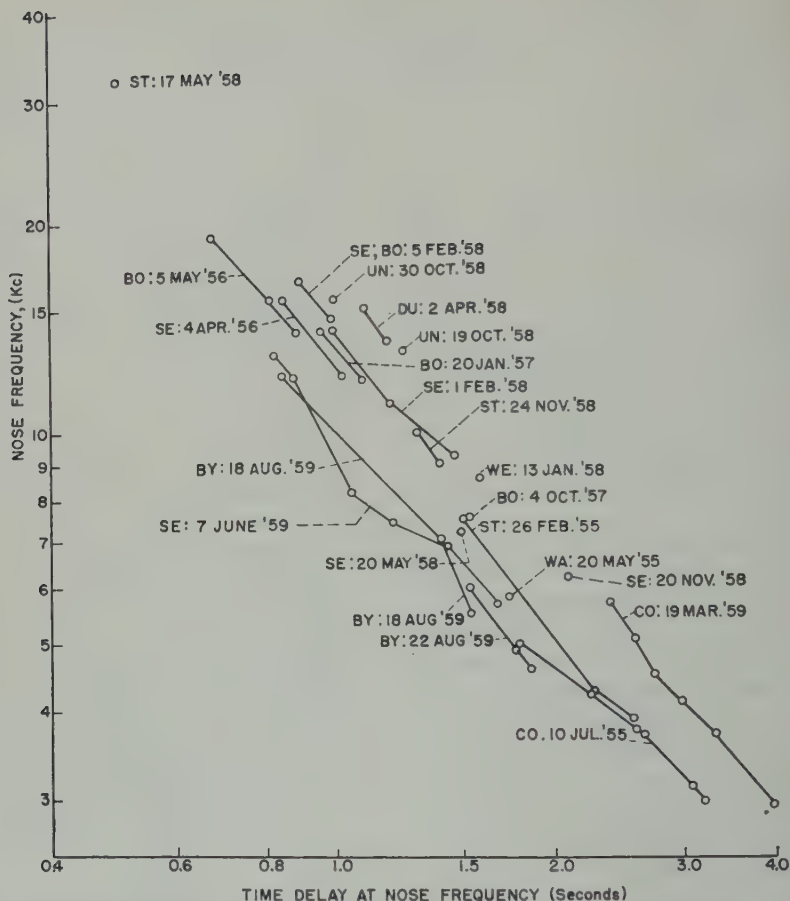


Fig. 2. Data showing the time delay and nose frequency of nose whistlers. *BO*, Boulder, Colorado; *BY*, Byrd Station, Antarctica; *CO*, College, Alaska; *DU*, Dunedin, New Zealand; *SE*, Seattle, Washington; *ST*, Stanford, California; *UN*, Unalaska, Alaska; *WA*, Washington, D. C.; *WE*, Wellington, New Zealand.

simple relation to the effective latitude of a received nose whistler component. One possible explanation is that whistler energy may, on occasion, travel for a considerable distance under the ionosphere. Another possibility is that the effective duct may terminate well above the lower ionosphere, thus allowing a small part of the total ray path to depart from a field line path. The lines that join some of the points usually indicate nose whistler groups, although they sometimes indicate a collection of different nose frequency-time delay points taken from whistlers occurring in the same day at the same station. One noticeable feature of the nose whistler groups is that the slopes of the lines are quite similar, even though the data were taken

at different times at different stations. The whistler group data thus indicate that the background ionization in the outer ionosphere usually has a reasonably smooth distribution with small enhancements of ionization to account for the discrete whistler components. Furthermore, the shape of the ionization distribution is more variable than the scale factor of the distribution.

The nose whistler data also strongly suggest a seasonal variation of time delay and background ionization density. The data for the months of May through August have consistently less time delay than the data for November through March. The time delay for the mid-year months is approximately the same during both sunspot

and maximum. The seasonal variation can be described in more detail after the model of ionization density is obtained.

Comparison of outer ionosphere models with data. The ionization distribution cannot be obtained directly from a single nose whistler trace. However, the nose whistler data can be compared with that predicted from various theoretical models.

Dungey [1954] proposes from thermodynamic arguments that collisions cause the density of particles in the outer ionosphere to be distributed in thermal equilibrium whether or not a magnetic field is present and even though collisions are very infrequent. Dungey's model, assuming only protons and electrons at a temperature of 1500°K are present, is

$$N = N_0 \exp (2.5/R) \quad (3)$$

where N_0 is the density at infinity and R is the distance measured in earth radii from the center of the earth. Johnson [1960] has recently derived a similar formula, in which he has included a small correction factor to account for centrifugal force. In the region of interest here the equations are nearly the same.

The presence of ducts of enhanced ionization suggests that the outer ionosphere does not exist in thermodynamic equilibrium. In the absence of collisions, charged particles tend to collect around a set of field lines, enclosing a constant amount of flux. There is, therefore, a tendency for the density of particles to be proportional to the magnetic field strength. We will call this model the 'gyrofrequency model' for convenience. Such a model has been proposed by Gallet [1959]. Storey had previously suggested this model, apparently because the regular low frequency dispersion integral can be directly integrated. Assuming a constant base level ionization, this model is given by

$$N = K \frac{f_H}{(1 + 3 \sin^2 \theta)^{1/2}} \quad (4)$$

where f_H = gyrofrequency.

θ = geomagnetic latitude.

Johnson [1959] and the author have independently suggested a combination model in which the ionization is proportional to the product of the magnetic field strength and an

exponential factor. Again, to keep the base level constant, we will use the formula

$$N = K \frac{f_H}{(1 + 3 \sin^2 \theta)^{1/2}} \exp (2.5/R) \quad (5)$$

Johnson has since rejected this model in favor of the one discussed previously.

The above three models were used to compute theoretical nose frequency-time delay curves. A small correction ($\Delta D = 5$) was applied to account for the additional dispersion in the F layer. The theoretical curves are compared to the nose whistler group data in Figure 3. (The factor $(1 + 3 \sin^2 \theta)^{-1/2}$ should be included in the labels for the curves in which the factor f_H appears.) A change in the scale factor for the theoretical models would shift the curves to the right or left, but the slopes would remain substantially unchanged.

The slope of the nose frequency-time delay characteristic for the Dungey model clearly does not fit the majority of the data. Each nose whistler trace can give only a weighted integral of plasma frequency along one set of field lines. The ionization is distributed much more rapidly along the field lines than across the field lines. One might ask if the Dungey model could fit the

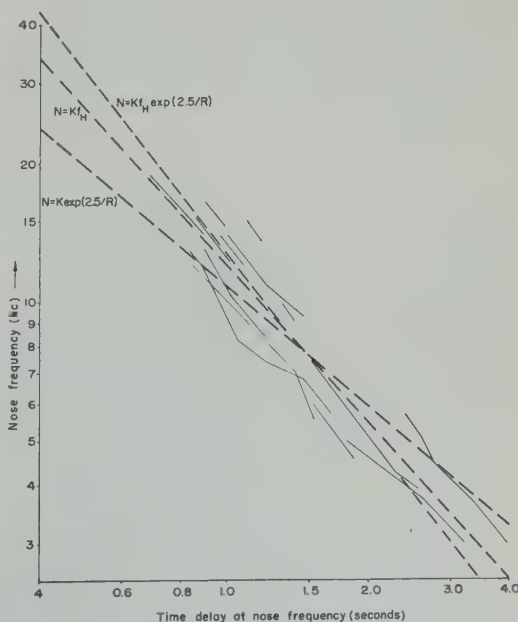


Fig. 3. Comparison of nose whistler data with the time delay as a function of frequency predicted from three models of the outer ionosphere.

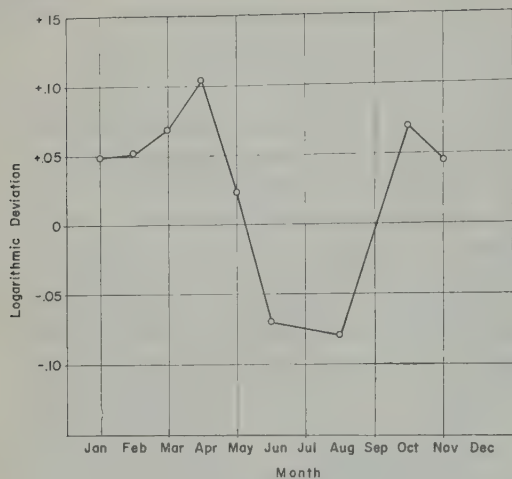


Fig. 4. Annual variation of nose frequency time delays.

whistler data if the base level of ionization were allowed to vary with latitude. The nose whistler data cover effective base latitudes from about 50° to 65° . Over this 15° range the base level would have to change by a factor of 4 or more to bring the model into reasonable agreement with the nose whistler data. There is no reason to expect such a large change in the base level.

The exponent in the Dungey model is proportional to atomic weight and inversely proportional to temperature. Some calculations were made using larger exponents for the Dungey model. The model

$$N = N_0 \exp(8/R) \quad (6)$$

fits the data quite well. However, the larger exponent would imply a temperature of 500°K if hydrogen is the main constituent in the outer ionosphere, or a temperature of 7000°K if nitrogen or oxygen were the main constituent. These temperatures differ considerably from temperatures found in satellite drag data (1250°K) [Harris and Jastrow, 1959] or helium escape calculations (1500°K) [Spitzer, 1948].

Of the two remaining models, the gyrofrequency model appears to fit the data somewhat better, both on the basis of the slope measurements and the over-all data, including the single nose whistler data points. An average model of the outer ionosphere density is then found to be

$$N = 12,000 f_H \quad (7)$$

where N is the electron density per cubic meter and f_H is the gyrofrequency in cycles per second.

Having obtained an average model we now estimate the annual variation in electron density. The monthly average logarithmic deviation of time delays to the nose are shown in Figure 4. These data cover the years 1957–1958 only. The nose whistler data for 1955–1956 are probably not statistically significant, but a nose whistler recorded on February 26, 1957, indicates that the seasonal variation during sunspot minimum is much smaller than during sunspot maximum. Supporting data, which indicate an annual variation in the ionization density, are the monthly dispersion averages shown in Figure 5 for the year 1958. The dispersion data alone would not be sufficient evidence since it could be argued that the variations in dispersion could result from seasonal movements in the ducts of enhanced ionization. Since the time delays at a given frequency and the dispersion values are proportional to the square root of electron density, the data in Figure 4 indicate an annual variation of electron density of 2:1. Possible factors which may be related to the annual variation are discussed by Helliwell [1961].

Other work. The first attempt to obtain electron distribution in the outer ionosphere from whistler data was made by Allcock [1956] without using nose whistlers. Average whistler time delays were aurally determined at a number of stations. The whistlers were assumed to travel along field lines terminating at the receiving station. By using these data and further assuming a radially symmetric distribution,

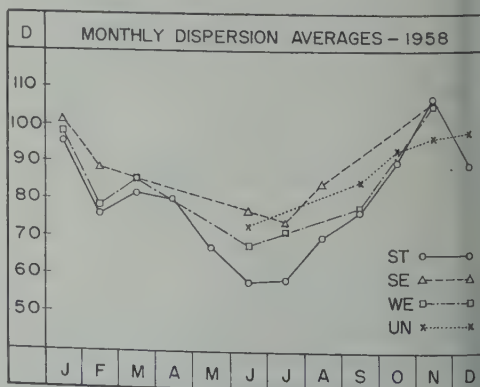


Fig. 5. Annual variation of whistler dispersion.

was able to derive an electron-density profile. The resulting densities are somewhat higher than those obtained in the present report. There are two principal reasons for the higher values. As we have seen, the geomagnetic latitude of a station does not necessarily represent the effective geomagnetic latitude of a received whistler. At a given station, whistlers can be expected to come from a range of effective latitudes around the station latitude. Since the whistler rate increases rapidly with latitude for the stations used by Alcock, the average whistler time delay at a station would correspond to the time delay at a lower geomagnetic latitude. Alcock discusses several sources of error, the nose whistler effect, but does not correct for it.

On the basis of records obtained at College Station on the day, March 19, 1959, Pope [1961] deduced that the ionization distribution is similar to that given in equation 5. The data from this day are included in Figures 2 and 3. It can be seen from the figures that the slope of nose frequency versus time delay for this day is not typical of the majority of data. It might be deduced that the gyrofrequency model would be consistent with the data from March 19, 1959, if the base level ionization were made to decrease slightly with latitude.

Variations of density across field lines. The whistler group recorded at Seattle on June 19, 1959, shows large deviations from a simple gyrofrequency model. No simple model could account for the irregular variation of nose frequency as a function of time delay. The time delays at nose frequencies of 5.65, 6.82, 11.9, and 13.0 kc are consistent with the simple gyrofrequency model. The time delays of the two components with nose frequencies of 7.35 and 11.9 kc are, however, less than the expected delay by approximately 18 per cent. This indicates a 40 per cent reduction in ionization density along the paths followed by these two nose whistler components. Thus large variations of ionization density may exist across the field lines. The variations noted here suggest that the effects of enhanced ionization of the size sufficient to trap whistler energy are easily obtained.

Results. The electron-density distribution deduced from the present study is shown in Figure 6. (This figure is the same as that shown in Figure 6 of Smith and Helliwell [1960].) The annual

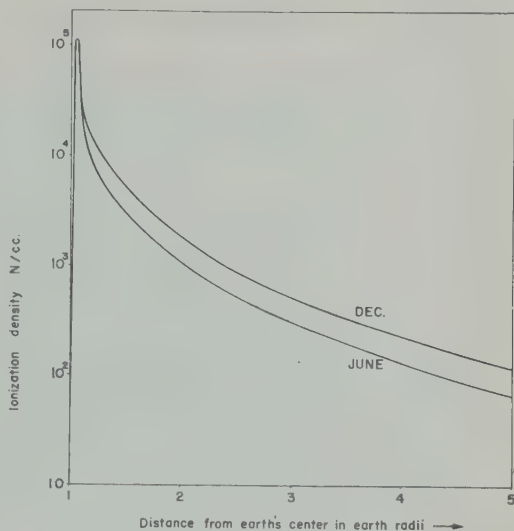


Fig. 6. Gyrofrequency model of the outer ionosphere.

variation is indicated by the separate curves for June and December. The distributions are fitted to a simple parabolic approximation of the lower ionosphere with a maximum plasma frequency of 3 Mc, corresponding roughly to the nighttime F_2 layer critical frequency. The average density at 5 earth radii, about 100 per cc, corresponds very closely to the density deduced by Obayashi [1958] from geomagnetic fluctuation data. However the interpretation of these data has since been refuted by Jacobs [1960].

The densities deduced from whistlers are most accurate near the top of the field line path. The present data thus apply best to the region from $2\frac{1}{2}$ to 5 earth radii. By comparing the data with other models we can deduce that the gyrofrequency model gives the correct density at any place in this region within a factor of 2. The uncertainty in the integrated ionization in this region should be much less.

From recent rocket flights, Berning [1960] has measured ionization densities to 1500 km. He deduces an ionization density of 10^4 /cc at this height, but indicates that the true value may be as low as 5×10^3 /cc. The nose whistler data, using the gyrofrequency model, gives values between 7.8×10^3 and 1.5×10^4 /cc. It is worth mentioning that reasonably accurate values of ionization densities to heights of 2000 km, deduced perhaps by rocket or satellite measurements, will greatly increase the accuracy

of whistler measurements at far greater distances.

The present data indicate a large annual variation in electron density in the outer ionosphere during sunspot maximum. There is insufficient data to determine the variations near sunspot minimum. An adequate amount of data will probably be collected during the coming sunspot minimum to answer this question. Present indications from nose whistler and dispersion data indicate that the annual variations will decrease in amplitude in the next few years.

The data also show that the ionization density does not vary in an entirely smooth and simple manner with position in the outer ionosphere. The ducts of enhanced ionization have lifetimes of only a few hours. Attempts to explain the distribution of ionization in the outer ionosphere must be based on dynamic considerations.

Acknowledgments. I wish to express my appreciation to Dr. R. A. Helliwell for his guidance and stimulating counsel during the course of this work. I am grateful to the many people and agencies who have supplied data for this project. The bulk of the data was obtained from the Whistlers-West project during IGY and IGC by the National Science Foundation grants Y-6.10/20 and NSF-G 8839, and cooperating agencies. Pre-IGY data at Stanford were obtained under Air Force contract AF19(604)-795. Nose whistlers from Byrd Station, Antarctica, were analyzed under contract NSF-G 13465. Nose whistlers from the Geophysical Institute, College, Alaska, were supplied by J. Pope. Nose whistlers from Washington, D. C., were supplied by H. Dinger. Other useful data and help were supplied by G. McK. Allecock, L. R. O. Storey, R. Barrington, W. C. Johnson, Don Jones, J. Katsurakis, D. Carpenter, and J. Crary.

The research was sponsored in part by the Air Force Office of Scientific Research under contract AF18(603)-126.

REFERENCES

- Allecock, G. McK., The electron density distribution in the outer ionosphere derived from whistler data, *J. Atmos. Terrest. Phys.*, **14**, 1959.
- Berning, W. W., A sounding rocket measurement of electron densities to 1500 kilometers, *J. Geophys. Research*, **65**, 2589-2594, 1960.
- Carpenter, D. L., Identification of whistler sources on visual records and a method of routine whistler analysis, *Tech. Rept. no. 5*, contract AF18(603)-126, Radioscience Laboratory, Stanford University, March 15, 1960.
- Dungey, J. W., Electrodynamics of the outer atmosphere, *Sci. Rept. no. 69*, contract AF19(122)-144, Ionospheric Research Laboratory, Pennsylvania State University, September 1954.
- Gallet, R. M., The very low frequency emission generated in the ionosphere, *Proc. IRE*, **211-231**, 1959.
- Garriott, O. K., A note on whistler propagation in regions of very low electron density, *Geophys. Research*, **63**, 862-865, 1958.
- Harris, I., and R. Jastrow, An interim atmosphere derived from rocket and satellite data, *P. Space Sci.*, **1**, 20-26, 1959.
- Helliwell, R. A., Exospheric electron density variations deduced from whistlers, *Ann. géophys.*, **(1)**, 1961.
- Jacobs, J. A., as reported in *Trans. Am. Geol. Union*, **41**, p. 627, 1960.
- Johnson, F. S., The structure of the outer atmosphere including the ion distribution above the F_2 maximum, *Tech. Rept. no. LMSD-4*, Lockheed Missiles and Space Division, Sunnyvale, California, April 28, 1959.
- Johnson, F. S., The ion distribution above the F_2 maximum, *J. Geophys. Research*, **65**, 577-580, 1960.
- Obayashi, Y., Geomagnetic storms and the earth's outer atmosphere, *Rept. Ionospheric Research Japan*, **12**, 301-335, 1958.
- Pope, J. H., An estimate of electron density in the exosphere by means of nose whistlers, *Geophys. Research*, **66**, 67-75, 1961.
- Smith, E. J., Characteristics of the extraterrestrial current system: Explorer VI and Pioneer, *J. Geophys. Research*, **65**, 1858-1861, 1960.
- Smith, R. L., The use of nose whistlers in the study of the outer ionosphere, *Tech. Rept. no. 6*, contract AF18(603)-126, Radioscience Laboratory, AFOSR TN-60-861, July 11, 1960a.
- Smith, R. L., Guiding of whistlers in a homogeneous medium, *J. Research NBS*, **64D (5)**, 1960b.
- Smith, R. L., Propagation characteristics of whistlers trapped in field-aligned columns of enhanced ionization, *J. Geophys. Research*, **66**, 3699-3707, 1961.
- Smith, R. L., and R. A. Helliwell, Electron densities to 5 earth radii deduced from nose whistlers, *J. Geophys. Research*, **65**, 2583, 1960.
- Smith, R. L., R. A. Helliwell, and I. W. Yabroff, A theory of trapping of whistlers in field-aligned columns of enhanced ionization, *J. Geophys. Research*, **65**, 815-823, 1960.
- Spitzer, L., The terrestrial atmosphere above 100 km., Chapt. VII of *The Atmospheres of the Earth and Planets*, Chicago University Press, pp. 213-249, 1948.
- Storey, L. R. O., An investigation of whistlers, *Phil. Trans. Roy. Soc. London*, **A, 246**, 113-141, 1953.
- Storey, L. R. O., A method to interpret the dispersion curves of whistlers, RPL 23-4-1, Defence Research Board of Canada, Ottawa, Canada, April, 1958.
- Yabroff, I. W., Computation of whistler ray paths, AFOSR TN-60-71, December 1959.

(Manuscript received May 17, 1961; revised August 29, 1961.)

Concerning Radiosondes, Lag Constants, and Radio Refractive Index Profiles

B. R. BEAN AND E. J. DUTTON

*Central Radio Propagation Laboratory
Boulder Laboratories, Boulder, Colorado*

Abstract. Temperature and relative humidity reported by the radiosonde have long been converted to radio refractive index without regard to sensor time lag. Past studies of radio ducting incidence have shown that under restricted conditions only corrections for humidity sensor lag need be made. It is here shown that lag constants currently in the literature are inadequate under conditions of superrefraction of radio waves. Use of currently available lag constants indicates that, should any correction be made, both temperature and humidity sensor lags must be considered for any climatological comparison of ducting incidence. By ignoring sensor time lag, one tends to underestimate ducting incidence; by correcting only for humidity sensor lag, one tends to overestimate ducting incidence.

Introduction. For several years data obtained from radiosondes have been analyzed per se the distribution of the radio refractive index with respect to height. The radio refractive index, n , is usually given as

$$(1) 10^6 \equiv N = \frac{77.6}{T} \left(P + \frac{4810e_s f}{T} \right) \quad (1)$$

where N is the refractivity, P the total atmospheric pressure (millibars), f is the relative humidity expressed as a decimal, and e_s is the saturation water vapor pressure (millibars) at temperature T ($^{\circ}\text{K}$) [Smith and Weintraub, 1954].

In the determination of N from radiosonde data one is subject to all the errors inherent in the radio observation. Recently Wagner [1960] has analyzed the errors in N arising from time lag of the sensing elements, data transmission techniques, and significant level selection criteria. Of the various sources of error, Wagner concludes that the time lag of the sensing elements is the most serious source of error. Further, for the south-California coastal inversions he concludes that only the time lag of the humidity strip should be considered. A similar conclusion has been reached by Clarke [1960] for practical applications involving shipborne radar and over-the-air-to-air communications. Although there is no significant correction associated with the time lag of the United States radiosonde's lithium chloride (LiCl) humidity sensing element, there is also a time lag in the temperature ele-

ment, which, as will be shown, must also be taken into consideration. The correction for the temperature element yields a twofold correction to N due to the actual error in temperature and the ancillary correction in vapor pressure resulting from the more correct estimate of the true saturation vapor pressure. This arises from the fact that when the LiCl element measures relative humidity it must be used with the saturation vapor pressure to obtain the actual vapor pressure. Since the saturation vapor pressure of water is a function of temperature, an error in temperature produces an error in the estimated water vapor pressure.

The lag constant of the LiCl humidity element becomes significantly larger for temperatures below 0°C [Dunmore, 1938]. Wexler [1949] has made detailed studies of the lag constants of the LiCl elements at low temperatures under laboratory conditions. Bunker [1953], however, has found quite different lag constants in the free atmosphere. He attributes this discrepancy in lag constants to the laboratory-determined values, which were obtained for isothermal conditions, whereas, in rising through the free atmosphere, the radiosonde normally passes from warm to cooler air. Bunker has raised a serious point, namely, that the temperature lag of the LiCl element is possibly as important as the isothermal humidity lags studied by Wexler. It is quite possible that the interplay of these two lags could produce a total effect either greater or less than the humidity

lag alone. We are now in a quandary, since there is currently in the literature no complete analysis of the interplay of the temperature and humidity lags of the LiCl humidity element. Although Bunker treats some aspects of this problem, he does not consider the case of decreasing humidity and increasing temperature, a typical condition giving rise to the superrefraction of radio waves. The differences between Wexler's and Bunker's estimates of the LiCl lag constants are so great as to cast doubt on the validity of applying any correction for this effect. Since neither Wexler's nor Bunker's tabulation of lag constants is complete, the choice of lag constant appears to be arbitrary. Wexler's lag constants will be adopted for the present discussion. Although neglecting the temperature lag of the humidity element may well yield considerable error, the present paper will attempt to demonstrate that, should any correction be made to radiosonde data, both the temperature and humidity elements must be corrected for their respective time lags for the purpose of preparing refractive index profiles.

In what follows, the theory of sensor time lag will be examined. Data from several climatically diverse locations will then be examined to illustrate the relative importance of the various lag-constant corrections under conditions of superrefraction of radio waves.

Theory of sensor time lags. Middleton and Spilhaus [1953] give, as the basic differential equation of the time lag of a meteorological sensor, measuring the variable θ ,

$$d\theta_i/dt = -(1/\lambda)(\theta_i - \theta_e) \quad (2)$$

where t is time, λ the appropriate lag constant, and the subscripts i and e stand for the indicated and environmental values, respectively. The solution to (2) depends upon the manner in which the environmental value of θ varies. For example, if we assume that θ_e varies linearly with time,

$$\theta_e = \theta_0 + \beta t \quad (3)$$

we obtain for the solution to (2)

$$\theta_e - \theta_i = +\beta\lambda[1 - \exp(-t/\lambda)] \quad (4)$$

as compared with

$$\theta_e = \frac{(\beta\lambda - 1) \exp(t/\lambda)\theta_i}{\beta\lambda \exp(\beta t) - \exp(t/\lambda)} \quad (5)$$

under the assumption that θ_e varies exponentially with time,

$$\theta_e = \theta_0 \exp(-\beta t)$$

For a column of air, the initial reading of humidity strip, θ_0 , is normally known, and an assumed linear decrease of θ with height; coefficient β becomes

$$\beta = \frac{\theta_i - \theta_0}{t - \lambda[1 - \exp(-t/\lambda)]}$$

with the result that (3) is written

$$\theta_e = \theta_0 + \frac{(\theta_i - \theta_0)t}{t - \lambda[1 - \exp(-t/\lambda)]}$$

Once the value of λ is determined, estimates of the true humidities of the air can be found at all heights up to the point where the gradients change.

We may proceed by a different course, noting in (2) that

$$\frac{d\theta_i}{dt} = \frac{d\theta_i}{dh} \frac{dh}{dt} = R \frac{d\theta_i}{dh}$$

where R is ascension rate of the radiosonde (assumed a constant 300 m/min).

If it is further assumed that θ_i varies linearly between reported values (an assumption compatible with radiosonde reporting procedure) then

$$R \frac{d\theta_i}{dh} = R \frac{\theta_{i,k+1} - \theta_{i,k}}{h_{k+1} - h_k}$$

where the k th and the $(k+1)$ st layers are the boundaries considered. Thus the environmental value of θ can be determined from

$$\theta_{e,k+1} = \theta_{i,k+1} + R\lambda \frac{\theta_{i,k+1} - \theta_{i,k}}{h_{k+1} - h_k}$$

which involves, in a simple fashion, only indicated or actually measured values of parameter θ .

It is assumed that θ_i and θ_e are identical at $h=0$ and that successive application of (11) will yield a more realistic estimate of the distribution of θ with height. When the correction procedure of the form of equation 11 is applied to temperature, where the temperature lag constant, λ_T , is always 3 sec [Sion, 1955], the corrected temperature profile is obtained immediately. The same is true for humidity, provided that the temperature

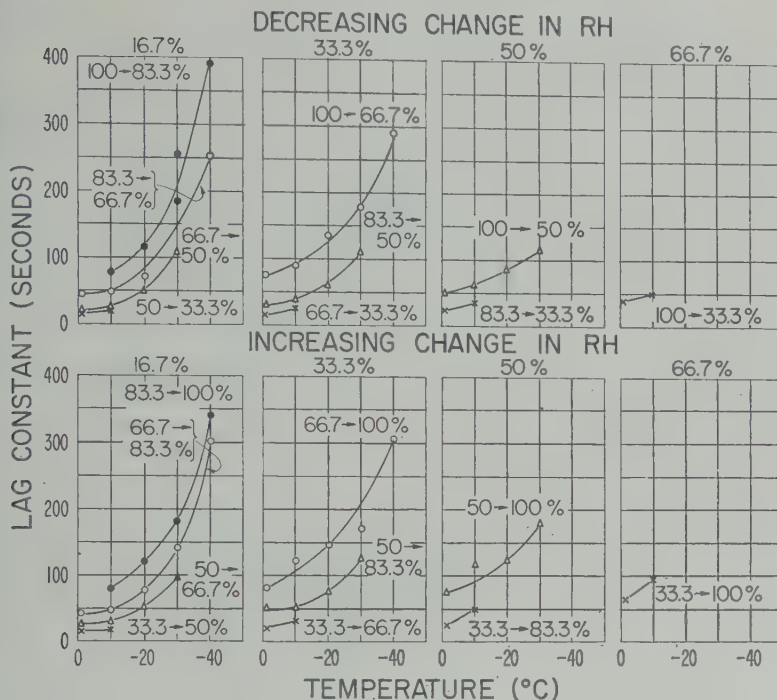


Fig. 1. Factors affecting the lag constant. After Wexler [1949].

than 0°C , when λ_r is assumed to be 10 seconds [Wagner, 1960], although is some indication [Wexler, Garfinkel, Hasegawa, and Krinsky, 1955] that for temperatures the humidity lag constant is nearer 5 seconds. For temperatures less than 0°C , however, λ_r is a function of true temperature, true value of f , and change of the value of f (see Fig. 1). This means that a relative solution must be used for f , since f changes as the estimate of the true value and Δf changes. Since our knowledge of λ_r is essentially empirical, the correction procedure is limited to the temperatures and values of Δf reported by Wexler.

Applying the above equations we generally find that the time lags are always known at the environmental and indicated values identical at the base of each layer [Wagner, 1960]. These two conditions are approximately satisfied only for the ground layer since the lag constant of the humidity strip is not known. It is not clear that any correction for lag may be made above the initial layer, or for subsequent layers, the initial indicated environmental values are not identical and,

further, lag constants have not been determined for this case.

Radiosonde profile analysis. The utility of the above lag-constant corrections is illustrated by analyzing past radiosonde data for the occurrence of ground-based radio ducts. A ground-based radio duct is one in which the gradient of N is sufficient to refract a radio ray to the same curvature as that of the earth. Thus, for ducting, since ray curvature is given by the gradient of the refractive index,

$$\frac{\Delta N}{\Delta h} \leq \frac{-1}{r_0} = -157 N \text{ units/km} \quad (12)$$

where r_0 is the earth's radius. The data analyzed were from the months of expected maximum duct occurrence at Fairbanks, Alaska (February), Washington, D. C. (May), and Swan Island, West Indies (August). All data were for the year 1953.

As an example of past work, Wagner [1960] has assumed that the refractivity lagged its environmental value according to (4) with the result that

$$N_e - N_i = \beta_N \lambda_N [1 - \exp(-t/\lambda_N)] \quad (13)$$

TABLE 1. Comparison of Ground-Based Ducting Statistics Derived from Various Sensor Time Corrections

Station	Reported Data			Corrected for T and f Sensor Lag		Assuming $\lambda_N \equiv \lambda_f = 10$ and Equation	
	No. Profiles	No. Ducts	Average Gradient, N units/km	No. Ducts	Average Gradient, N units/km	No. Ducts	Average Gradient, N units/km
Washington, D. C. (May), temperate climate	123	1	-163	9	-202	6	-15
Swan Island (August), tropical climate	62	15	-186	20	-206	23	-22
Fairbanks, Alaska (February), arctic climate	51	2	-211	4	-212	12	-25

After an analysis of the meteorological conditions of his area of application, he set $\lambda_f \equiv \lambda_r \equiv 10$ seconds; i.e., the time lag of N derived from pressure, temperature, and humidity was identically that of the humidity sensor for $T > 0^\circ\text{C}$. Comparing this method of N -lag correction with the uncorrected data (first and last columns of Table 1), we note an increase in both intensity and incidence of ducts in all climates. This is particularly marked for Fairbanks, where a sixfold incidence increase is obtained. However, when individual time-lag corrections are made for both temperature and humidity by means of (11) a quite different picture of duct statistics is obtained. For example, the Fairbanks data indicate a twofold rather than a sixfold increase. On the other hand, Washington shows an incidence of 9 rather than 6 ducts with a marked increase in N gradient when both temperature and humidity are corrected for. The change of ducting statistics at all three locations obtained by the two methods of time-lag correction yields paradoxical results. The near contradiction of the two correction procedures is easily explained and serves as an illustration of the necessity of correcting both the temperature and humidity elements for general application to ducting statistics.

Consider typical temperature and humidity conditions associated with ground-based ducts within each climate. Such cases are shown in Figure 2. It is sufficient to note that the temperate ducts arise from typical radiation inversion conditions of increasing temperature and decreasing relative humidity with height; the

arctic ducts are associated with the intense arctic radiation inversion with ground temperature near -25°C and nearly constant relative humidity with respect to height; the tropical ducts, however, appear to be due to small decreases of both temperature and humidity with height at temperatures near $+25^\circ\text{C}$. The effect of sensor lag upon these different gradients is always to make the indicated gradient appear less than the true or environmental gradient. Thus correcting for sensor lags makes the temperature and humidity gradients more intense. This, in turn, affects the resultant N gradient. We may write

$$dN = \frac{\partial N}{\partial T} dT + \frac{\partial N}{\partial e} de + \frac{\partial N}{\partial P} dP$$

which, for normal conditions of 15°C , 1013 mb, and 60 per cent relative humidity, is approximated by

$$\Delta N = -1.27 \Delta T + 4.50 \Delta e + 0.27 \Delta P$$

Since, for ducting, a large negative gradient of N is required, the effect of temperature sensor lag correction is to make temperature increases with height more pronounced, with a resultant larger contribution to a negative N gradient. Humidity corrections under all conditions lead to a large decrease of relative humidity with height, as indicated and concomitant more rapid decrease of N with height. Such an explanation does not seem so evident for tropical conditions.

ver, where both temperature and relative humidity decrease with height. This apparent box is explained by the relatively great change in saturation vapor pressure of water associated with a small change of temperature of 25°C , which then produces the required decrease of e , and N , with height. It is then, that the temperature sensor lag correction produces an added humidity correction. Therefore, both sensing elements must be corrected for the purpose of preparing N profiles. In the case of the arctic inversion the interplay of the temperature and humidity lag corrections is very important, as has been noted by Berg [1950].

Thus, the correction for humidity sensor lag

alone tends to *overestimate* ducting incidence since the corrected relative humidity decrease, coupled with the *indicated* temperature increase, produces a greater decrease in water vapor pressure than is actually present.

Conclusion. It appears from the present study that, if studies of refractive index profile characteristics are to include sensor lag correction, allowance should be made for both temperature and humidity sensor time lags, regardless of climate. Any systematic application of these conclusions to large bodies of data, however, must await the appearance in the literature of effective lag coefficients that combine the effects of both temperature and humidity lags upon the LiCl element.

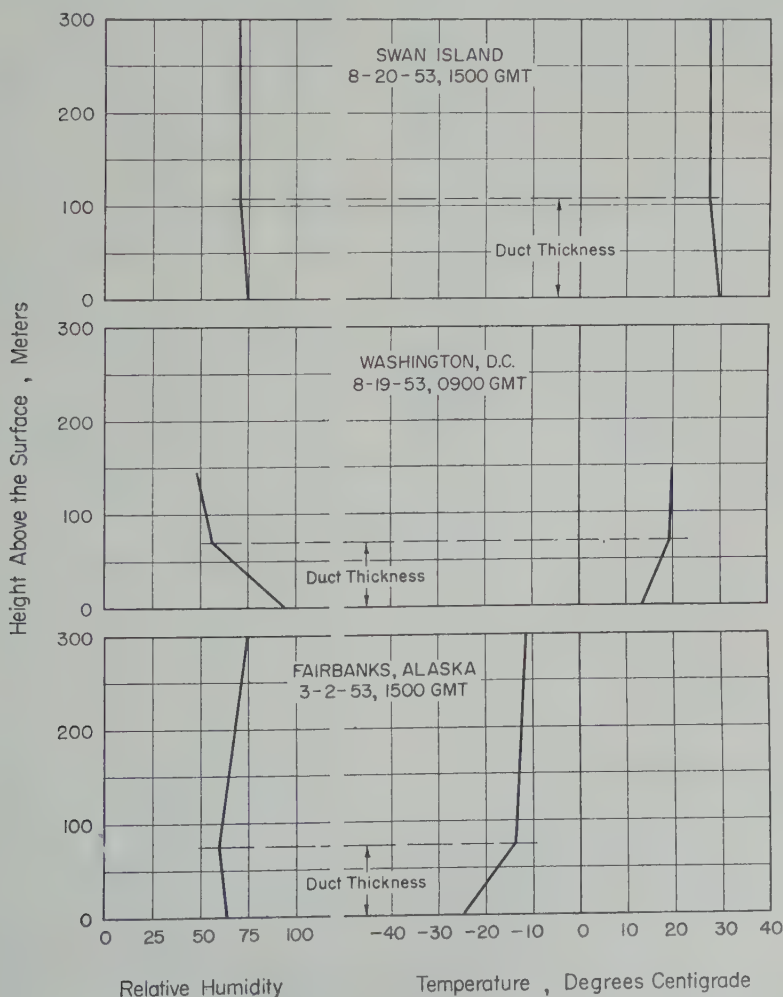


Fig. 2. Temperature and humidity profiles for typical ground-based ducts.

Acknowledgments. We express our gratitude to the readers for their clarifying comments on the two humidity element lag constants.

This work was sponsored in part by Task 31 of the Navy Weather Research Facility, Norfolk, Virginia.

REFERENCES

- Bunker, Andrew F., On the determination of moisture gradients from radiosonde records, *Bull. Am. Meteorol. Soc.*, **34**, 406-409, 1953.
- Clarke, L. C., Theory of atmospheric refraction, part II of *Meteorological Aspects of Radio-Radar Propagation*, pp. 31-82, 1960.
- Dunmore, Francis W., An electric hygrometer and its application to radio meteorography, *J. Research NBS*, **20**, 723-744, 1938.
- Middleton, W. E. K., and A. F. Spilhaus, *Meteorological Instruments*, p. 63, University of Toronto Press, Toronto, 1953.
- Sion, Elio, Time constants of radiosonde thermistors, *Bull. Am. Meteorol. Soc.*, **36**, 16-21, 1955.
- Smith, E. K., Jr., and S. Weintraub, Constants in the equation for atmospheric refractive index at radio frequencies, *J. Research NBS*, **50**, 3, 1953.
- Wagner, N. K., An analysis of radiosonde errors on the measured frequency of occurrence of ducting layers, *J. Geophys. Research*, **65**, 2085, 1960.
- Wagner, N. K., The effect of time constants of radiosonde sensors on the measurement of temperature and humidity discontinuities in the atmosphere, *Bull. Am. Meteorol. Soc.*, **42**, 317, 1961.
- Wexler, Arnold, Low temperature performance of radiosonde electric hygrometer elements, *J. Research NBS*, **43**, 1949.
- Wexler, A., S. Garfinkel, F. Jones, S. Hasegawa, and A. Krinsky, A fast responding electric hygrometer, *J. Research NBS*, **55**, 71-78, 1955.
- Yerg, Donald G., The importance of water vapor in microwave propagation at temperatures low freezing, *Bull. Am. Meteorol. Soc.*, **31**, 177, 1950.

(Manuscript received August 10, 1961.)

The Solar Semidiurnal Atmospheric Oscillation

K. A. SMALL AND S. T. BUTLER

*The Daily Telegraph Theoretical Department
School of Physics, University of Sydney
Sydney, Australia*

Abstract. It has been found that ozone heating is capable of producing atmospheric 'tidal' oscillations whose semidiurnal component is in close agreement with ground-level observations, both in magnitude and in phase. No assumption of an atmospheric resonance is required.

Solar semidiurnal tide in the atmosphere, will be referred to as S_2 , has an amplitude at ground level of about 1.5 mb, which is a magnification of about 100 over gravitational equilibrium tide; the maxima occur around 10 A.M. and 10 P.M. local time. In spite of this large magnification, previous calculations [Pekeris, 1937; Jacchia and Kopal, 1952; and Sen and White, 1955] have considered atmospheres which, on account of an atmospheric resonance, yielded large, purely gravitational contributions.

But the existence of free periods of oscillation in the atmosphere depends sensitively on the vertical temperature distribution, in particular, the oscillation with a period of half a solar day depends critically on the temperature maximum assumed in the 50-km region. This maximum must be in the vicinity of 300°K [Jacchia and Kopal, 1952], which can be reconciled with recent measurements of the atmospheric temperature profile [Murgulla, 1957]. These put an upper limit to the temperature of 300°K, and consequently we are led to the conclusion that as far as S_2 is concerned the atmosphere is decidedly nonresonant and that the gravitational contribution to atmospheric tides must be a relatively minor one. A difficulty immediately posed is that previous work, e.g. by Wilkes [1951] and Sen and White [1955], has indicated that thermal source contributions are approximately of equal magnitude and therefore, in the absence of resonance, are too small to account for the observations.

Here, we wish to report here results of a calculation of the thermally forced oscillations

in a nonresonant atmosphere due to the periodic heating of the ozone layer, a thermal contribution not previously evaluated; the results obtained for the S_2 tide at ground level agree closely with observation both in magnitude and in phase.

In setting up the calculation the assumption is made that the net energy gained through ozone absorption in the ultra-violet and visible regions of the direct solar beam during the hours of sunlight is lost at a constant rate over the entire day through emission in the infrared by O_3 and the other constituents of the atmosphere. This assumption, namely, of a mean radiative equilibrium, would certainly be invalid in the troposphere, but here the forcing function is effectively zero; the method should, however, be sensibly applicable in the stratosphere where are found the maxima of both the ozone density (~ 25 km), and its absorption (40–50 km).

According to White [1956] the magnification of S_2 over the equilibrium tide for both gravitational and thermal excitation is

$$A(x) = 1 + N + i \delta e^{x/2} \left\{ N \left[\left(\frac{H}{h} - \frac{1}{2} \right) + \frac{d}{dx} \right] I(x) \right]_g - \left(\frac{d}{dx} - \frac{1}{2} \right) I(x) \right\} \quad (1)$$

where the subscript g refers to ground level, and $\delta = \text{constant} = 7.5 \times 10^8 \text{ sec}$.
 $H = \text{scale height}$.

$$x = \int_0^z \frac{dz}{H};$$

$z = \text{height above sea level}$.

h = eigenvalue of atmosphere corresponding to $S_2 = 7.9$ km.

N = purely gravitational contribution.

The function $I(x)$ is derived from the thermal forcing term under the assumption of variable separability. If $Q(z, \theta, \varphi, t)$ is the heat absorbed per unit mass per unit time at (z, θ, φ) , where θ = colatitude, φ = longitude, t = world time, then, for S_2 ,

$$Q = q(z)\Theta_2^2(\theta)e^{i(\sigma t + 2\varphi)}$$

where $2\pi/\sigma$ is the 12-hour period, and $\Theta_2^2(\theta)$ is a solution of Laplace's tidal equation.

Then $I(x)$ is a particular solution of the radial equation [Sen and White, 1955] given by

$$I(x) = y_1(x) \int_{-\infty}^x \frac{y_2(x)T(x) dx}{y_1' y_2 - y_2' y_1} - y_2(x) \int_{-\infty}^x \frac{y_1(x)T(x) dx}{y_1' y_2 - y_2' y_1}$$

where

$$T(x) = (Rqe^{-x/2})/gC_v\gamma^2 h$$

and γ = ratio of specific heats (C_p/C_v). Also y_1 and y_2 are independent solutions of the reduced radial equation.

In the determination of $q(z)$ the dependence of the ozone distribution on latitude and time of year was established from the fact that the total amount of ozone in a vertical column has (1) a minimum at the equator and a maximum around 60° latitude, (2) a maximum in early spring and a minimum in the fall, (3) an annual variation that increases with latitude.

From the data supplied by Craig [1947] we make the following fit to the ozone distribution:

$$\begin{aligned} \rho_{O_3}(z, \epsilon, t') &= [\rho_{O_3}(z)]_{\text{equator}} \\ &\times \{20.25\epsilon^2(1 - 2|\epsilon|) + 1\} \\ &\times \left(1 + \frac{\epsilon}{2} \cos \frac{2\pi t'}{T}\right) \end{aligned}$$

¹ Various authors [Craig, 1947] have postulated circulations as a possible explanation of the latitudinal dependence of the ozone distribution; as a result there would be an equatorial maximum in the ozone distribution above, say, 35 km, while below 35 km there is an equatorial minimum with maxima at 60° latitude. We note here that the inclusion of this effect in the present analysis would tend to slightly increase the magnitude of the semidiurnal tide because of the form of $\Theta_2^2(\theta)$.

where $\epsilon = \frac{1}{2} - \theta/\pi$, and t' = time measured from the northern spring equinox with $T = 1$ year.

As a first example we have computed the function $q(z)$ for the sun moving northward from the equator (northern spring equinox). The result can be expressed in the form

$$q(z) = (0.939\alpha_2(z) - 0.177\alpha_4(z))$$

where

$$\alpha_{2n} = \frac{1}{2\pi} \frac{(2n-2)!}{(2n+2)!} \int_0^\pi d\theta$$

$$\cdot \int_{-\pi}^{+\pi} d\varphi E(z, \theta, \varphi) P_{2n}^2(\cos \theta) \sin \theta$$

and

$$\begin{aligned} E(z, \theta, \varphi) &= \frac{\rho_{O_3}}{\rho_{\text{air}}} \int d\lambda k_\lambda I_{O_\lambda} \\ &\cdot \exp\left(-k_\lambda \int_0^s \rho_{O_3}(s) ds\right) \end{aligned}$$

$\rho_{O_3}, \rho_{\text{air}}$ = densities of O_3 and air, respectively

k_λ = absorption coefficient O_3 per unit path length at NTP.

I_{O_λ} = solar intensity outside atmosphere at λ

$\int d\lambda$ = integration over the absorption spectrum

s = path length in atmosphere of the ozone from the sun to the point (z, θ, φ)

The temperature profile used is given in Table 1, and our computations yielded the following result:

$$1 + N = 1.41e^{-12\epsilon^2}$$

$$I_{x=0} = 1.5534 \times 10^{-7} \text{ sec}^{-1}$$

From equation 1 we find for the ground amplification $A = A(x=0)$

$$A \approx 74e^{96\epsilon^2}$$

We see here that in a non-resonant atmosphere, with a purely gravitational amplification of S_2 of only 1.4, the heating of the lower layer is capable of producing a semidiurnal oscillation at ground level of an amplitude that is close to the actual observed for S_2 . As a result it appears that tidal oscillations in the lower atmosphere, that is below 35 km, are due mainly to the heating of the lower layer.

Siebert [1961] estimated the semidiurnal

TABLE 1. Mean Vertical Temperature Distribution
Temperature profile [Murgatroyd, 1957] used in calculation of ozone thermal heating.

T, °K	z, km	T, °K	z, km	T, °K
288	40	270	80	200
271	42	279	82	201
254	44	288	84	204
239	46	295	86	208
227	48	299	88	212
220	50	300	90	217
217.5	52	299	92	221
217	54	297	94	226
218	56	290	96	232
219	58	282	98	237
220	60	273	100	243
221	62	263	102	249
223	64	253	104	255
226	66	243	106	262
229	68	234	108	268
233	70	225	110	275
238	72	218	112	282
245	74	210	114	290
252	76	205	116	298
261	78	202	118	306

is excited by the solar heating of the layer, concluding that it was in no way compatible with the observed tidal amplitude. Contrary to our present conclusions, but it can be shown that this discrepancy stems from the unrealistic, though tractable, temperature profile Siebert used. Indeed, our calculations also lead to a much smaller effect

when we insert the Siebert temperature profile. The results of our present work therefore stress the importance of using a realistic profile in close agreement with observation.

A full description of this work will be published in a later paper. Further work is also in progress on the altitude dependence of the ozone heating and the amplitudes of other tidal components, together with the seasonal variations of the ozone layer.

Acknowledgements. All the numerical computations were performed on Silliac, the digital computer of the University of Sydney, School of Physics. One of us (K. A. Small) held a Sydney University research studentship during the time this work was performed. Finally, the authors express their indebtedness to Professor H. Messel for providing the excellent research facilities available in the School of Physics.

REFERENCES

Craig, R. A., *Meteorol. Monographs*, 1, 1947.
Jacchia, L. G., and Z. Kopal, *J. Meteorol.*, 9, 13, 1952.
Murgatroyd, R. J., *Quart. J. Roy. Meteorol. Soc.*, 83, 417, 1957.
Pekeris, C. L., *Proc. Roy. Soc. London, A*, 158, 650, 1937.
Sen, H. K., and M. L. White, *J. Geophys. Research*, 60, 483, 1955.
Siebert, M., *Advances in Geophys.*, 7, 105, 1961.
White, M. L., *J. Geophys. Research*, 61, 489, 1956.
Wilkes, M. V., *Proc. Roy. Soc. London, A*, 207, 358, 1951.

(Manuscript received July 11, 1961;
revised August 14, 1961.)

Traveling Pressure Waves Associated with Geomagnetic Activity

PETER CHRZANOWSKI, GARY GREENE, K. T. LEMMON, AND J. M. YOUNG

National Bureau of Standards
Washington 25, D. C.

Abstract. Traveling atmospheric pressure waves with periods from 20 to 80 seconds and pressure amplitude from about 1 to 8 dynes/cm² have been recorded at a microphone station at Washington, D. C., during intervals of high geomagnetic activity. Trains of these waves can be expected at Washington from a quadrant approximately centered on north whenever the magnetic index K_p rises to a value above 5. Their horizontal phase velocity across the station is usually higher than the local speed of sound. During two 'red' auroras, clearly visible at Washington and at lower latitudes, the 20- to 80-second-period waves were accompanied by longer period, higher pressure, and much slower traveling pressure disturbances. Observational data on the wave systems are presented and discussed.

Introduction. The records from an array of microphones located near Washington, D. C., recently indicate the occurrence of traveling pressure waves in the atmosphere associated with geomagnetic activity. Reports of audible sounds near auroral displays have been discussed by Chapman and Bartels [1940] and others. The infrasonic, or below audible, sounds discussed here have predominant periods in the 20 to 80-second range with very occasional periods up to 300-second waves.

It is the purpose of this paper to report some of the distinguishing characteristics of the magnetic storm signals, and to pose questions about the nature and origin of the disturbance and relationships to other electromagnetic phenomena.

Technique. The infrasonic waves are detected by a system of four microphones placed in a quadrilateral averaging 7.5 km on a side (Fig. 1). The output of each microphone is transmitted to the laboratory, where it is passed through amplifiers with various passbands as indicated by the periods at the output points for the different channels (Fig. 1). Response curves for the amplifiers used in this study are shown in Figure 2. The amplifier outputs are recorded on translucent paper tapes with speeds appropriate to the passbands of the different channels. The presence of a traveling wave was established when the same waveforms could be found on all four records with a common passband and when time displacements between the four microphone sites were con-

sistent with the passage of a plane wavefront. Time differences between microphone sites were used to compute the direction from which the wavefronts arrived and the trace velocity or horizontal phase velocity of the waves. Experience has shown that many traveling waves can be found by this procedure that would not be distinguishable on a single record from background pressure variations due to wind turbulence or other causes.

February 11, 1958, storm. The infrasonic

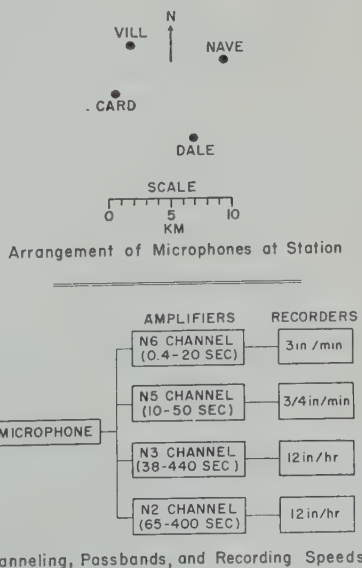


Fig. 1. Schematic diagram of recording equipment showing the relative positions of the microphones and the different frequency ranges used.

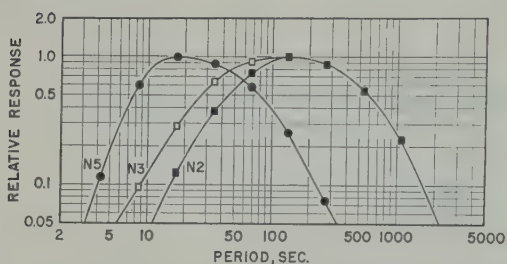


Fig. 2. Relative response curves of infrasonic systems used in this study.

waves recorded during the magnetic storm of February 11, 1958, exhibited a more or less typical behavior pattern. The storm began suddenly on February 11 at 0126 UT. It was ac-

companied by an intense red aurora visible far south as Florida. Within the first 3-h interval on February 11 the planetary magnetic index K_p rose to 9. Photographs of the records from equipment with the 10- to 50-cm bandpass, showing the onset of the acoustic signal, are displayed at the top of Figure 3. The two lower strips are photographs of pencil tracings of the superposed waveforms aligned to show the sound waves and long-period slow-traveling waves, respectively. The latter two are discussed below. The first distinguished sound waves arrived about 0642 UT from an azimuth of 340° with respect to geographic north with a horizontal phase velocity of 7 m/sec. The waves decreased in amplitude to

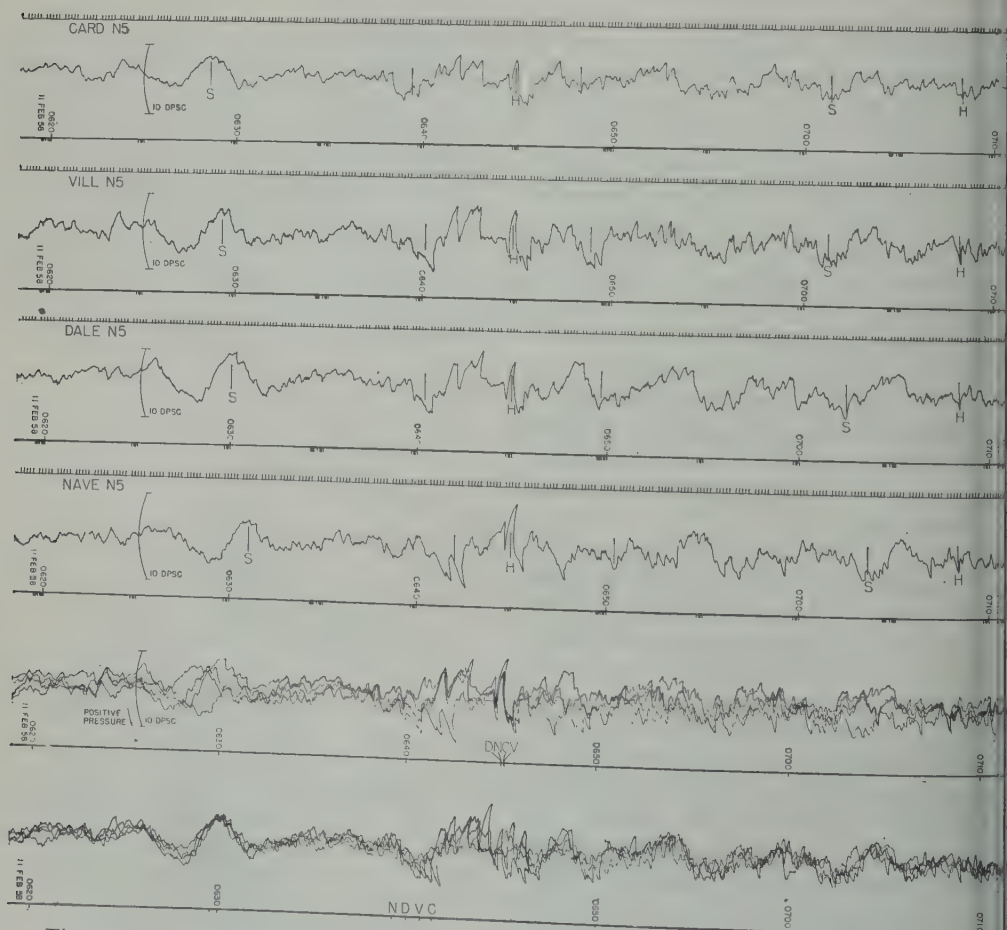


Fig. 3. N5 records from four microphones during early part of infrasonic signal on February 11, 1958. The two bottom strips are pencil tracings with corresponding waveforms overlaid, showing short transit times corresponding to acoustic waves apparently incident at a high angle with the plane of the microphones and long-period slow-traveling waves, respectively.

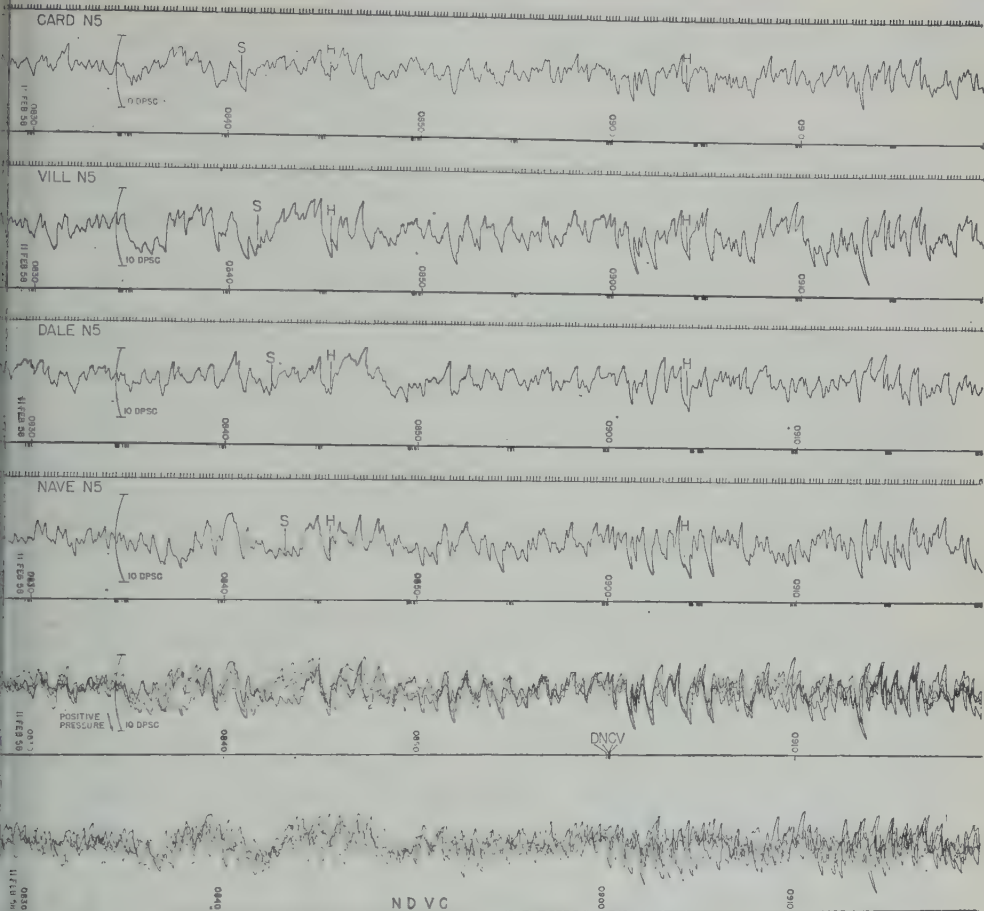


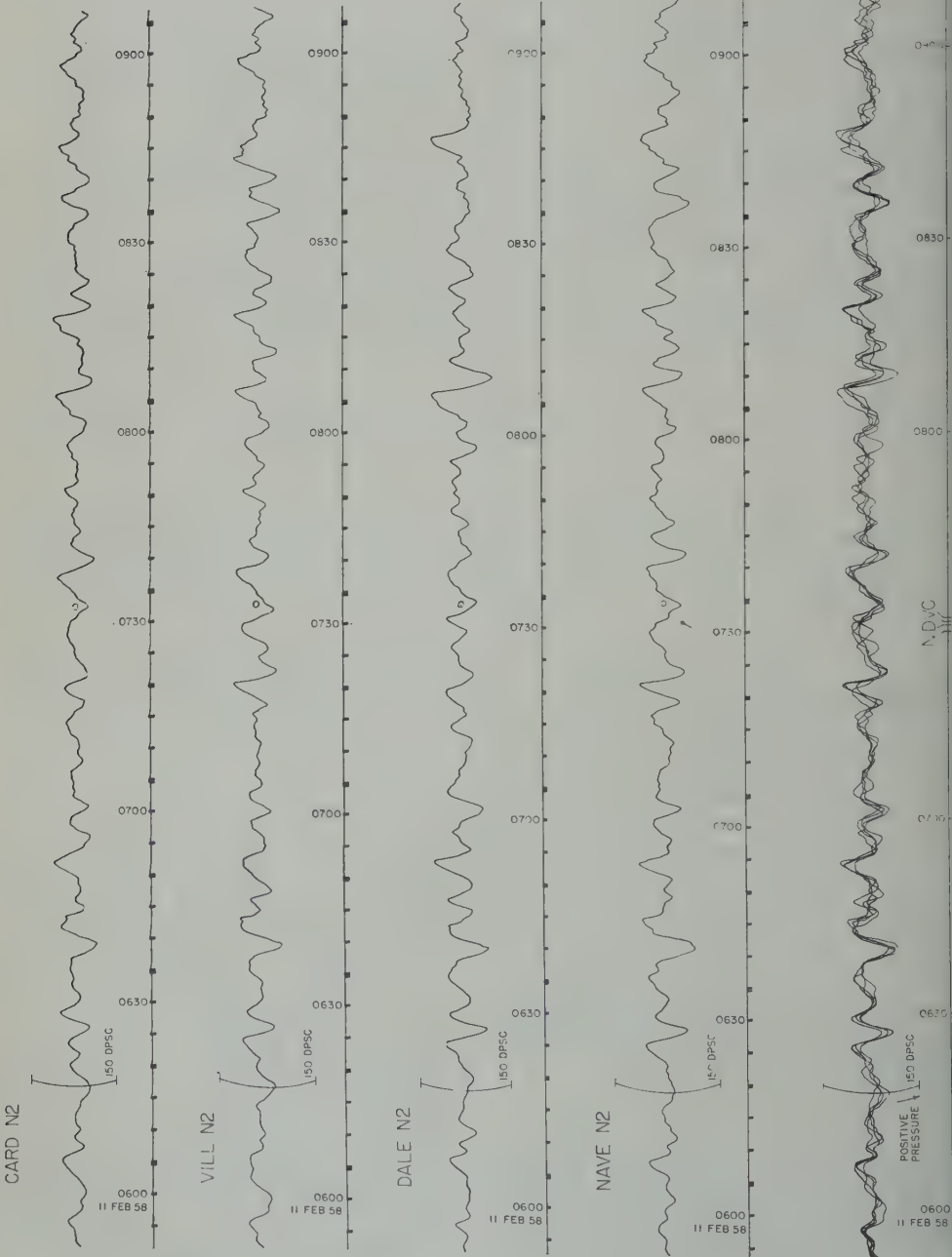
Fig. 4. N5 records during the main part of the signal on February 11, 1958.

when a second burst of waves arrived from with a velocity of 550 m/sec. At 0905, s similar to those at 0642 arrived from 332° a trace velocity of 750 m/sec. These bursts shown in Figure 4. The sound waves ded in amplitude gradually and disappeared een 1100 and 1200 UT. No further indica- of sound waves were found on the records veral days.

Winkler, Peterson, Hoffman, and Arnoldy], working with balloon-borne equipment inneapolis, observed two bursts of auroral ys during the storm of February 11, 1958. first burst began at 0622 UT, and a second, intense burst occurred at 0850. Winkler d an apparent time relationship between X-ray bursts, magnetic bays, and cosmic absorption. Troitskaya [1961] reports ir- ar micropulsations with diminishing periods

(IPDP) coincident with the X-ray bursts. Sound bursts with very high trace velocity arrived at Washington, D. C., about 15 to 20 minutes after the bursts of auroral X rays were observed. A similar connection between times of occurrence of sound waves and X rays was found during the storm of July 15-17, 1959.

In addition to the sound waves, long-period slow-traveling pressure waves were recorded during February 11, 1958, which are evident in Figures 3 and 4. Records of these waves from the 65- to 400-second passband equipment in the 0600 to 0900 UT interval are shown in Figure 5. Periods in this interval averaged about 6 minutes with pressures about 40 dynes/cm². The waves subsided about 1415 UT and did not reappear until February 12 at 0300. Waves of this type are often present during intervals of magnetic activity, but they also



very frequently at other times. Periods from 3 to 15 minutes or more with trace velocities from 10 to 100 m/sec and typical pressures up to 100 dynes/cm². The direction of these waves corresponds in general to the direction of 500-mb winds over Washington, D. C. Similar waves have been observed by [Flauraud, Mears, Crowley, and Crary, 1957], and a similar correspondence with upper-level winds has been noted. No direct connection between these long-period waves and geomagnetic activity has been established.

Experimental procedure. The 3-hour range planetary magnetic index K_p was chosen as a measure of the degree of disturbance of the environment of the earth by solar corpuscular emissions. For an explanation of the meaning and method of derivation of K_p , the interested reader is referred to the articles of Bartels [1957]. The degree of association between sound waves and magnetic disturbances during which K_p rose to higher is shown in Table 1.

A second signal was considered to be associated with a magnetic disturbance if it occurred within a few hours of a rise of K_p to a value of 6 or higher and if the trace velocity of the wave was higher than the local speed of sound. Over the 19-month interval covered by the tabulation, there were 62 magnetic disturbances. During 16 of these disturbances the high background noise level at the station precluded finding a detectable pressure signal. On 10 occasions the background was considered to be reasonably favorable but no indications of sound waves were found. Not counting the occasions when the station was noisy, sound was present during 46 of 46 magnetic disturbances with K_p of 6 or higher, or for 78 per cent of the favorable circumstances. The degree of association is higher for the higher K_p indices.

An obvious question is how the number of sound signals occurring during geomagnetic disturbances compares with the number of signals occurring in the absence of magnetic activity. During the interval from June 1958 through December 1959, 92 signals were entered in the laboratory log. Thirty-six, or 39 per cent, occurred after a rise of K_p to 6 or higher. Thirteen of the remaining 56 occurred after magnetic disturbances for which the planetary index rose to a value of 5 only. Counting these also, 53 per cent of the signals are traceable to geomagnetic

TABLE 1. Sound Associated with Geomagnetic Activity Period from June 1958 to December 1959

Planetary Magnetic Index	Num- ber of Dis- turb- ances	Sound Present	Sound Absent	Sta- tion Noisy	Per- cent- age Associ- ation
9	4	4	0	0	100
8 or higher	13	10	0	3	100
7 or higher	28	20	1	7	95
6 or higher	62	36	10	16	78

activity. Most of the others are probably due to other natural phenomena. Among the possibilities are earthquakes (P. Chrzanowski and R. K. Cook, manuscript in preparation), volcanoes [Passechnik, 1958], and tornadic storms (P. Chrzanowski, J. M. Young, and H. L. Marrett, manuscript in preparation).

The histogram in Figure 6 relates the number of signals to their time of onset. Only signals occurring after magnetic disturbances during which K_p changed by at least 2 to a value of 6 or higher within two adjacent 3-hour intervals were considered. The sound waves show a tendency to appear roughly 3 to 6 hours after a sharp rise in the magnetic activity.

An estimate of the onset times of sound waves during 56 magnetically disturbed intervals is displayed in Figure 7. The histogram indicates that there is an interval between 0800 and 1400 local time when the beginning of sound waves is infrequent. A number of signals beginning before the interval were found to extend through it.

The variation of direction from which sound arrives with time of day is shown in Figure 8. The general trend is from the northeast in the

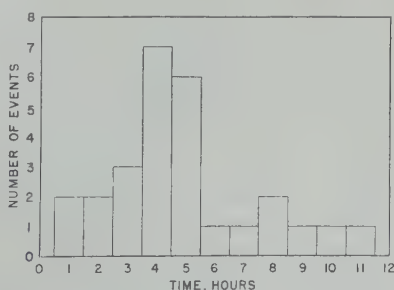


Fig. 6. Elapsed time between sharp rise in K_p and onset of sound waves.

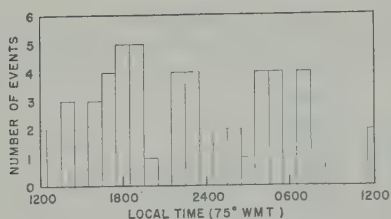


Fig. 7. Diurnal variation of onset times of sound waves associated with magnetic storms.

evening, through north about midnight, then northwest in the morning, shifting to the northeast again somewhat suddenly near local noon. The progress during four individual storms is shown in Figure 9. To avoid possible confusion with sounds from other sources, the data in Figures 8 and 9 were restricted to signals with trace velocities above 390 m/sec. The changes are not at all regular. One of the characteristics of these signals appears to be a certain amount of confusion due to almost simultaneous arrival of quasi-sinusoidal pulses from slightly different directions with different trace velocities.

Pressure amplitudes are typically 1 to 3 dynes/cm² and occasionally higher. During the magnetic storm on July 15, 1959, a pressure amplitude of 7 dynes/cm² and a period of 240 seconds was recorded. Sound disturbances usually last several hours. Occasionally a continual signal is present for as long as 24 hours. Onset and ending are often gradual and difficult to determine because of the noise background in the same spectral region.

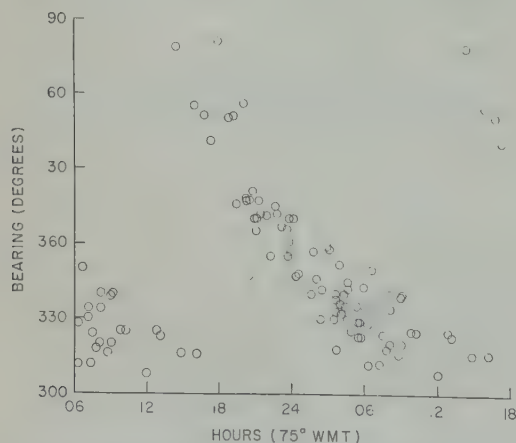


Fig. 8. Diurnal variation of direction of arrival of sound waves during magnetic storms.

Discussion. The high trace velocity along earth's surface, usually between 400 and 500 m/sec, and the absence of higher frequencies make it appear likely that these waves are generated in the upper atmosphere and approach the ground at a higher angle to the surface. For a distant source, a ray picture would indicate that the waves reach regions of the atmosphere where the temperature is of the order of 400° to 1500°K, and these temperatures occur above the 100-km level [Minzner, Champion, and Pond, 1959]. At heights near the *E* regions, high frequencies would be damped.

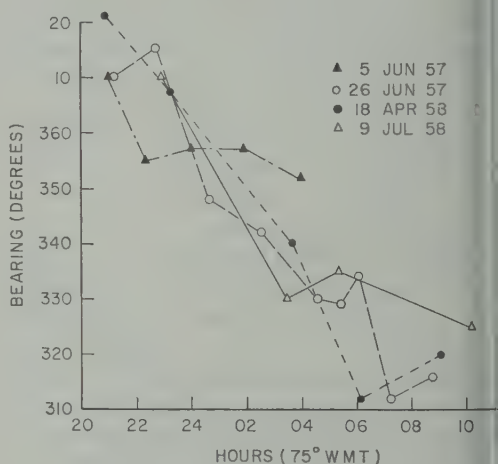


Fig. 9. Diurnal variation of direction of arrival of sound waves during specific magnetic storms.

One possibility is that the principal disturbance is trapped in the *E* layer and that what is observed is leakage from the bottom of the duct. The characteristic impedance seems to indicate that the disturbance in the *E* layer is associated with pressure fluctuations of 1 per cent or more. In this case nonlinear behavior may be important.

Observations from spaced stations should aid in determination of the source location and direction as well as propagation parameters of the disturbance in the atmosphere. The possibility that the source moves around near the auroral zones and always near geomagnetic midnight has been considered. There are other phenomena that suggest a somewhat localized source. Ellis [1960] has reported similar behavior in onset times of 5-kc/s electromagnetic noise in Australia during magnetic storms. He finds a lack of occurrence of on-

the same local time interval (150° EMT) as the gap in sound wave onsets indicated here. In addition, Ellis [1961] has reported two examples of 5-kc/s noise storms during which amplitude fluctuations on spaced stations correlate with a time difference approximately equal to the local time difference between stations. It is possible that the sources may be related. Jacchia [1959] has reported density fluctuations inferred from satellite drag during magnetic activity. Auroral motions determined from sky camera photographs [Kim and Currie, 1960; Bhattacharyya, 1960] show speeds comparable to the speed of sound and are predominantly westward in the early hours of the night and eastward toward morning. Similar motions have been inferred from radio reflections [Kaiser, 1961]. Fluctuations of intensity of auroral emission lines in the same period range of the present study were reported by Campbell and Rees [1961]. Corpuscular heating in the auroral zones at the mirror points of the outer Van Allen belt [Van Allen, McIlwain, and Ludwig, 1959; Jastrow, 1960] represents a possible source of pressure waves in the atmosphere. Many irregularities and irregular motions detected by present techniques were discussed by Hines [1960] in relation to possible interpretation in terms of atmospheric waves. A detailed study of the sources of these ionospheric signals and the energy released by them, in conjunction with electromagnetic measurements, should be very helpful in explaining the behavior of the earth's atmosphere during magnetic storms.

Acknowledgment. We wish to thank the members of the Infrasonics Group of the National Bureau of Standards and other colleagues whose contributions in equipment design and operation and assistance made this work possible.

REFERENCES

- Bartels, J., II, The technique of scaling indices K and Q of geomagnetic activity; III, The geomagnetic measures for the time-variations of solar corpuscular radiation, described for use in correlation studies in other geophysical fields, *IGY Ann.*, 4, 215-236, London, 1957.
- Bhattacharyya, B. K., A study of auroral motions from all-sky camera records, *Can. J. Phys.*, 38, 1279, 1960.
- Campbell, Wallace H., and M. H. Rees, A study of auroral coruscations, *J. Geophys. Research*, 66, 41, 1961.
- Chapman, S., and Julius Bartels, *Geomagnetism*, vol. 1, p. 466, Oxford University Press, 1940.
- Ellis, G. R. A., Geomagnetic disturbances and 5 kc/s electromagnetic radiation, *J. Geophys. Research*, 65, 1705-1710, 1960.
- Ellis, G. R. A., Spaced observations of the low-frequency radiation from the earth's upper atmosphere, *J. Geophys. Research*, 66, 19-23, 1961.
- Flauraud, E. C., A. H. Mears, F. A. Crowley, Jr., and A. P. Crary, Investigation of microbarometric oscillations in eastern Massachusetts, *Geophysical Research Paper* 27, Geophysical Research Directorate, Air Force Cambridge Research Center, ARDC, May 1954.
- Hines, C. O., Internal atmospheric gravity waves at ionospheric heights, *Can. J. Phys.*, 38, 1441-1481, 1960.
- Jacchia, Luigi G., Corpuscular radiation and the acceleration of artificial satellites, *Nature*, 183, 1662, 1959.
- Jastrow, Robert, Geophysical effects of the trapped partial layer, *Revs. Modern Phys.*, 32, 947, 1960.
- Kaiser, T. R., Radio investigations of aurora and related phenomena, *The Airglow and the Aurora*, edited by E. B. Armstrong and A. Dalgarno, Pergamon Press, New York, 1955.
- Kim, J. A., and B. W. Currie, Further observations of the horizontal movements of aurora, *Can. J. Phys.*, 38, 1366, 1960.
- Minzner, R. A., K. S. W. Champion, and H. L. Pond, The ARDC model atmosphere, 1959, *Air Force Surveys in Geophysics*, no. 115, Air Force Cambridge Research Center, Bedford, Massachusetts, 1959.
- Passechnik, I. P., Seismic and air waves which arose during an eruption of the volcano Bezymyanny, on March 30, 1956, *Bull. (Izv.) Acad. Sci. USSR, Geophys. Ser.*, no. 9, September 1958.
- Troitskaya, V. A., Pulsations of the earth's electromagnetic field with periods of 1 to 15 seconds and their connection with phenomena in the high atmosphere, *J. Geophys. Research*, 66, 5-18, 1961.
- Van Allen, J. A., C. E. McIlwain, and G. H. Ludwig, Radiation observations with satellite 1958, *J. Geophys. Research*, 64, 271-286, 1959.
- Winkler, J. R., L. Peterson, R. Hoffman, and R. Arnoldy, Auroral X rays, cosmic rays, and related phenomena during the storm of February 10-11, 1958, *J. Geophys. Research*, 64, 597-610, 1959.

(Manuscript received August 8, 1961.)

Measurements of Current Density in the Fair Weather Atmosphere

J. H. KRAAKEVIK

*U. S. Naval Research Laboratory, Washington 25, D. C.¹
and University of Maryland, College Park, Maryland²*

Abstract. Simultaneous measurements of potential gradient and conductivity have been made from an airplane over an altitude range of from 15 m to 6 km. Within the exchange layer over the oceans, the conduction current density decreases with altitude, requiring a convection current which is probably caused by the upward diffusion of positive space charge. The conduction current is a maximum at 15 m above the earth's surface, where it is found to be on the average some 40 per cent, and on some soundings 200 per cent, higher than that above the exchange layer. Overland, space charge convection produces a conduction current within the exchange layer which is randomly variable, and on the average some 20 per cent higher than that above the layer. A convection current of some 50 per cent of the total current is sometimes observed at the top of the exchange layer. Above it, the current density is constant with altitude to within 10 per cent, and is found to vary from 1.1×10^{-12} amp/m² over Maryland to 1×10^{-13} amp/m² over Greenland, approximately inverse to the variation in columnar resistance. The average value of current density over the oceans is 2.7×10^{-12} amp/m², giving an estimated world-wide conduction current of 1400 amp. Over Greenland, the total potential and average current density aloft on individual soundings both vary together with time and are in phase with the universal variation in thunderstorm activity. A comparison of simultaneous measurements at the surface and aloft in Greenland and near Key West, Florida, show that surface measurements of current density at a single station cannot be used to provide an index of world-wide effects on a day-to-day basis.

Introduction. After the earth was found to possess a negative charge during fair weather the atmosphere was discovered to be slightly positively charged, it was deduced that a conduction current flows in the atmosphere carrying positive charge downward. The average world-wide conduction current has been estimated at 1800 amp by *Gish* [1951], based on ocean-surface measurements of conduction current density. The current is generated by world-wide thunderstorm activity [*Gish and Wait*, 1950; *Stergis, and Kangas*, 1957b], which also causes a 10 per cent diurnal variation about this mean. *Gish* [1951] also estimated the total resistance between the earth and high atmosphere to be 200 ohms from free-air conductivity measurements. Assuming Ohm's law to hold, he estimated the potential of the high atmosphere, which also varies diurnally, to be, on the average, 60 kv positive with respect to the earth.

The fair weather atmosphere is usually considered to be in a state of quasi-static equilibrium and thus adjusts continually to slowly varying processes, such as the universal diurnal variation in current and potential. Hence the conduction current density has been considered vertical and constant with altitude. However, the only known simultaneous series of measurements of conductivity and potential gradient [*Everling and Wigand*, 1921] show that the conduction current density decreases with altitude.

Because of the contamination usually found within the surface exchange (austausch) layer, time averages of many days of surface measurements are required to extract universal effects from the larger local variations in atmospheric electric variables. It has been suggested by *Gish* [1951] that current density measured at the surface in selected locations might possibly provide a continuous index of world-wide thunderstorm activity, since it is less dependent upon local conditions than either conductivity or potential gradient. However *Holzer* [1955] and *Kasemir* [1957] observed purely local effects in potential gradient and conduction current density at sunrise and attributed them to the onset

¹Permanent address: Physics Department, Wheaton College, Wheaton, Illinois.
²Part of this work was submitted to the University of Maryland in partial fulfillment of the requirements for the Doctor of Philosophy degree in Physics.

of turbulence. Mühleisen [1958], basing his suggestion on laboratory measurements, considered that the 'sunrise effect' might be due to space charge produced by water evaporation at the earth's surface. The increase in conduction current is believed to be due to space charge convection, and its magnitude is estimated by Gish [1949] to be less than 30 per cent of the conduction current.

In this paper results are presented of simultaneous conductivity and potential gradient measurements from an instrumented airplane in order to show the importance of convection currents within the exchange layer. It is further shown that the assumption of quasi-static equilibrium is generally valid only above the exchange layer and that surface measurements of current density are not representative in general of those at altitude.

2. Method of measurement. A multiengine airplane was instrumented to measure both polar conductivities, the potential gradient, and pressure, temperature, and relative humidity. The instruments and measuring techniques have been described elsewhere [Kraakevik and Clark, 1958]. The total conductivity λ is obtained from the measured positive (λ_+) and negative (λ_-) polar conductivities, where

$$\lambda = \lambda_+ + \lambda_- \quad (1)$$

The ambient electric field intensity is vertical and negative (downward) in fair weather. It is related to the potential gradient by

$$E = -\partial V / \partial Z \quad (2)$$

The conduction current density j is computed from the measured electric field and conductivity, using Ohm's law, where

$$j = \lambda E \quad (3)$$

j also being negative in fair weather. The space charge density ρ is obtained from the altitude variation in potential gradient using Poisson's equation, where

$$\rho = -\epsilon(\partial^2 V / \partial Z^2) \quad (4)$$

and ϵ is the atmospheric permittivity.

Altitude distributions of the measured variables were obtained on vertical soundings between 15 m above the earth's surface and 6 km above sea level. Data were taken in steps on the

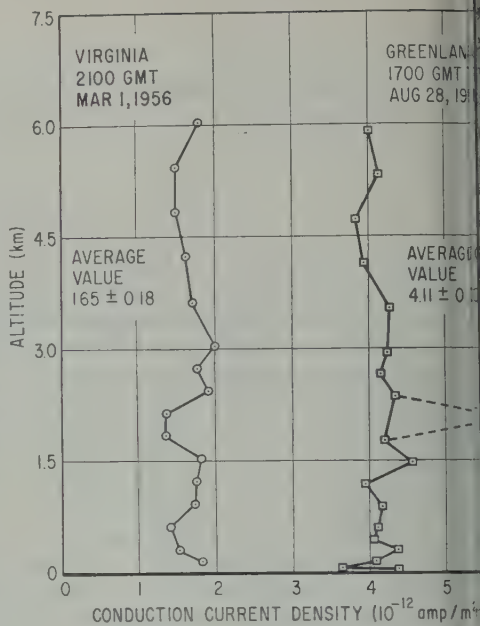


Fig. 1. Conduction current density distributions above Greenland fjord and Virginia plain.

descent, the time of the sounding being the quarter hour nearest the midpoint of the 70-minute interval required.

The investigation was restricted to fair weather conditions, in which no thunderstorms or precipitation was within 50 km and cloud cover was 10 per cent or less. Of 29 soundings attempted between March 1955 and May 1956, 22 successful fair weather soundings were made over a number of land and ocean areas between Bermuda and the California coast, and between Greenland and the Gulf of Mexico.

3. Altitude variations. Two typical altitude distributions or profiles of conduction current density are shown in Figure 1. One is an on-land profile obtained over Virginia on March 1, 1956, at 2100 GMT. The current density exhibits no tendency to increase or decrease with altitude and has an average value of 1.65×10^{-12} amp/m² with a standard deviation of 0.1×10^{-12} amp/m², or 10.8 per cent. Since the soundings were all taken during fair weather, the negative sign of j will be omitted when there is no ambiguity. The other profile was obtained over Søndre Strømfjord, Greenland, on August 28, 1955, at 1700 GMT. The average for this file is $4.11 \pm 0.18 \times 10^{-12}$ amp/m², or a standard deviation of 4.4 per cent. The difference

average values of j is due primarily to the difference in columnar resistance (resistance of $\pi r^2 l$ in cross section from the surface to a given height) at the two locations. The columnar resistance up to 6 km over Virginia is 1.1×10^{10} ohm-m² whereas over Greenland it is 6.1×10^{10} ohm-m².

The constancy of conduction current density with altitude shown in Figure 1 verifies that the atmosphere was approximately in quasi-static equilibrium during these soundings. It may be noted that for the March 1 profile the conductivity increased by a factor of 20 from the lowest altitude (15 m) to the highest (6 km), as the potential gradient decreased by the same factor, j being maintained nearly constant. The rather large standard deviation of 18 per cent is indicative of the general inhomogeneity prevalent in this region (eastern United States) resulting in horizontal and vertical inhomogeneities in the atmospheric electric field.

In the Greenland profile the conductivity increased exponentially with altitude [Kraakevik, 1956] by a factor of about 6 from 30 m to 6 km, indicating a very clean atmosphere electrically. Hence the conduction current density varies less relative to the mean value than in the March 1 sounding—only 4.4 per cent. The current anomaly of 5.9 units shown by the solid lines at 2.0 km is due to a lack of simultaneity in gradient and conductivity measurements. It was discovered in the analysis that the current data were taken when the airplane was in a thinly scattered cloud layer, some 2 min before the conductivity was measured, resulting in an abnormally high gradient. The conductivity data were taken outside the cloud layer. This result emphasizes the importance of truly simultaneous measurements of atmospheric electrical quantities.

The effect of convection on conduction current density distributions is illustrated in Figure 2. The curve on the left obtained near Bermuda on March 30, 1956, is representative of Atlantic results. For this sounding the average value of j above the surface exchange layer is 1.61×10^{-12} amp/m² ± 7.2 per cent. The high value of j on this flight is 3.27 units, occurring at an altitude of 15 m. This represents an increase of 103 per cent over the mean value. If the total current density be j_t , then

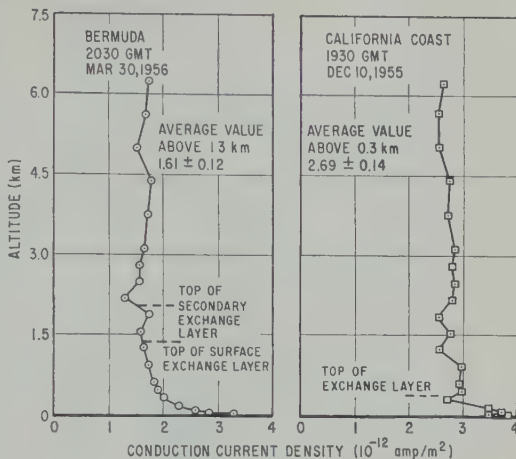


Fig. 2. Conduction current density distributions over the Atlantic and Pacific oceans.

$$j_t = j_c + j \quad (5)$$

where j_c is the convection current density. Note that in fair weather both j and j_t are negative (positive ions moving downward) whereas j_c is in the opposite direction (upward). Let it be assumed that the total current density j_t remains constant with altitude and is equal to the average conduction current density \bar{j} above the exchange layer. Equating \bar{j} above the exchange layer to j_t is considered justified because of the constancy of j with altitude, and the known reduction in turbulent convection, above the exchange layer. Then according to (5) the convection current near the earth's surface is $+1.66 \times 10^{-12}$ amp/m², or 103 per cent of the total current. This is considerably greater than the maximum of 30 per cent estimated by Gish [1949]. The values of j within the surface exchange layer are not included in the determinations of \bar{j} because of the influence of convection.

A corresponding profile made 20 miles off the California coast near San Francisco on December 10, 1955, is shown at the right of Figure 2. An average conduction current of 2.69×10^{-12} amp/m² ± 5.1 per cent is obtained above the exchange layer. The maximum value of j for this sounding again occurs near the earth's surface, and a magnitude of 3.84 units at 15 m is observed. This represents a convection current density of $+1.15$ units, or 43 per cent of the total current, under the assumptions given above. Several other profiles of current density which

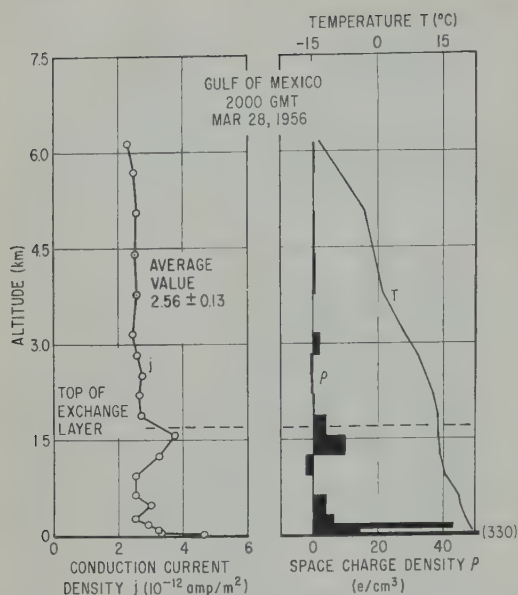


Fig. 3. Space charge convection at top of exchange layer.

illustrate convection effects are given in a previous paper [Kraakevik, 1958b].

The convection currents deduced from Figure 2 are in the opposite direction to the normal negative conduction current. Thus j_c requires the mechanical transport upward of positive space charge. The convection current density may be represented by

$$j_c = w\rho \quad (6)$$

as is done by Schonland [1953], where w is the vertical (positive upward) wind velocity. Equation 6 is valid for considering local changes in current density. However, it is probably not suitable for studying the average convection current, since the mean vertical wind velocity in fair weather is too small to support appreciable transport of space charge.

A more appropriate expression for convection current may be obtained by analogy with the theory of turbulent motion given by Sutton [1953]. This expression is, according to Gish [1949],

$$j_c = -A(\partial\rho/\partial Z) \quad (7)$$

where A is called the coefficient of eddy diffusion, corresponding to Sutton's exchange coefficients. The coefficient A is not constant in general but varies with altitude, as

$$A = A_1(Z/Z_1)^g$$

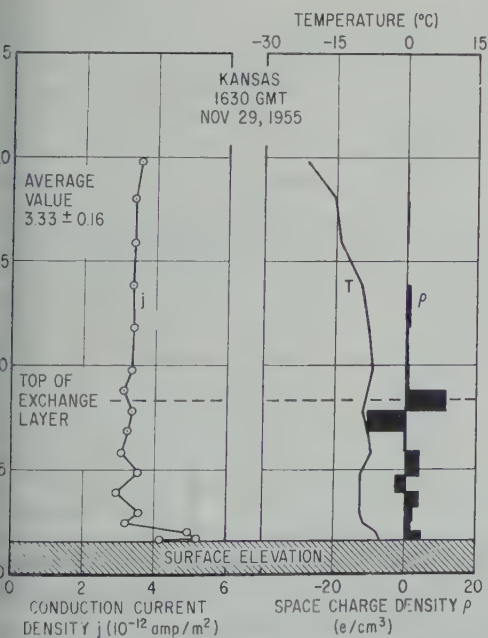
where A_1 is the value of A at Z_1 , usually $Z_1 = 1$ m, and g is an index of atmospheric stability. An expression similar to (8) is given by Tversky and Timofeev [1948] in a theoretical treatment of the problem of turbulent diffusion of space charge near the earth's surface. Convection current results are considered analytically in the next section.

The effect of the exchange layer on the convection current density is shown in Figure 3. This sound was made over the Gulf of Mexico near West, Florida, on March 28, 1955. The convection current density profile is shown at the top of the exchange layer and has an average magnitude of $2.56 \pm 0.13 \times 10^{-12}$ amp/m², or a spread of 5.2 per cent. This average excludes the three points below 0.1 km and the two points between 1.1 and 1.5 km, which are obviously influenced by convection. The maximum value of j_c is found at the lowest altitude reached (15 m), and it has a magnitude 81 per cent of the total current density. Positive space charge accumulates at the top of the exchange layer, corresponding to the top of the temperature inversion as shown by O'Brien [1958], because of the lower conductivity within the layer. This may be seen from the space charge and temperature profiles at the right of Figure 3. The narrow region of high charge density near the surface (off scale at 330 e/cm³) is probably due to the electrode effect.

In addition to space charge, occasional occurrence of turbulence at the top of the exchange layer is indicated by the presence of the large horizontal inhomogeneities in nucleus density in conductivity, and in water vapor at this level reported by Sagalyn and Faucher [1956]. The combination of a space charge gradient and turbulence produces convection current. This effect is observed on the March 28 profile, where a maximum j_c associated with the exchange layer top of 46 per cent of the total current density is seen. The correlation of regions of high conduction current density and large charge density can be seen in Figure 3. It should be noted that the space charge densities indicated are averages over the altitude intervals shown, but the actual charge densities may be much higher with narrower altitude increments.

That the presence of a large space charge density and a steep gradient in space charge

necessary but not sufficient conditions for convection current may be seen by examining the conduction current and space charge distributions for the sounding of November 29, 1955, at Kansas (Fig. 4). Excluding the three low-points (the ground elevation is 0.47 km for this location), which are uncertain because of the probable triboelectric charging effect in the conductivity due to dust particles, and also because of convection, the average magnitude of j is 3.33×10^{-12} amp/m² \pm 4.9 per cent. The base of the temperature inversion at 2.4 km corresponds approximately to the top of the exchange layer, the altitude difference between measurements being the increment of uncertainty. A large positive space charge is observed at this altitude, with a large negative space charge just below it, as reported by Clark [1958]. However, although space charge is present, there is insufficient vertical motion of the air to produce an observable convection current. The stability of the atmosphere is indicated by the temperature profile in Figure 4, which shows the atmosphere to be practically isothermal from the surface up to a 3.5-km altitude. Thus from Figures 3 and 4 it is shown that the large space charge density (about 10 e/cm³



4. Space charge without convection at top of exchange layer.

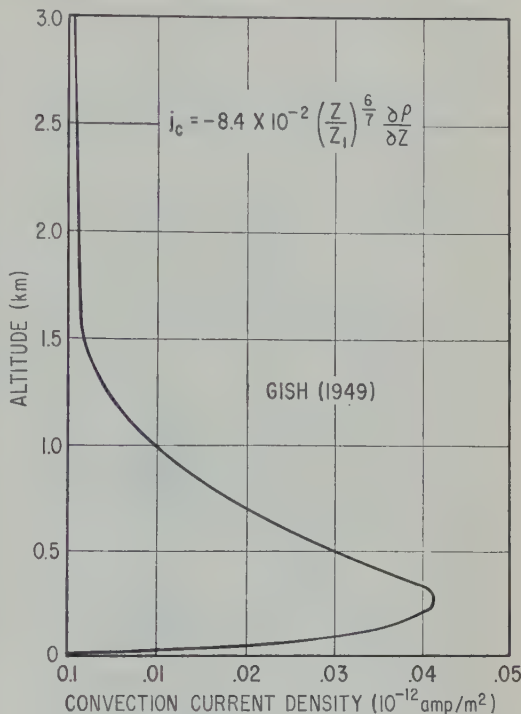


Fig. 5. Convection current density profile estimated by Gish [1949].

or more), which changes with altitude, and turbulence (indicated by instability or a steep lapse rate) are required conditions for observable convection currents.

4. *Convection current.* No known attempt has been made previous to this investigation to determine experimentally the magnitude and variation with altitude of the convection current density in fair weather. Perhaps the best presently available estimate, based on Schweidler's potential gradient results over Europe and on equations 7 and 8, is given by Gish [1949] and shown in Figure 5. From values for A_1 and g suggested by Gish, and values for $\partial \rho / \partial Z$ obtained by differentiating Gish's empirical expression for space charge density, convection current may be obtained as a function of altitude. It may be seen from Figure 5 that the maximum value of j_c is approximately 0.04×10^{-12} amp/m² and that it occurs at an altitude of 0.25 km. This maximum is less than 2 per cent of the average conduction current over the oceans. It should be noted that the scale in Figure 5 is expanded 100 times over that normally used for current density. Gish further

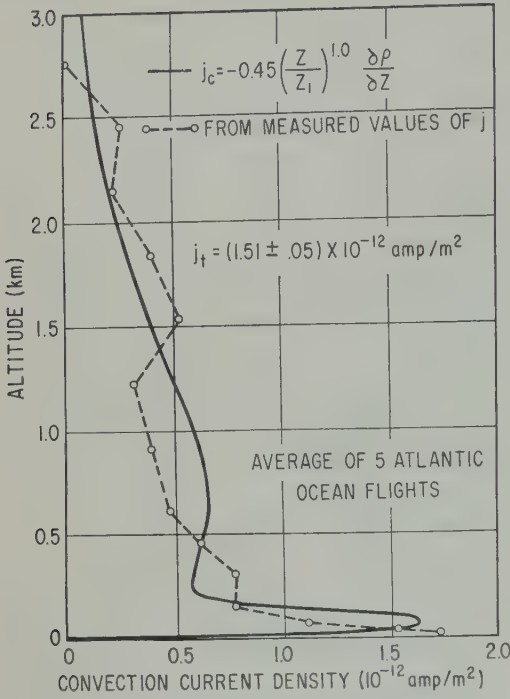


Fig. 6. Average convection current density profile off east coast of the United States.

stated that calculations based on Brown's space charge observations in California indicate that, although $\partial\rho/\partial Z$ is 38 times the maximum value used in the figure, the convection current at 1 m is less than 1 per cent of the average ocean value of j . That these estimates are too low is shown in the preceding section.

An analysis has been made of the convection current distributions for five profiles off the east coast of the United States. These flights, made between 1530 and 2030 GMT, include three profiles obtained 150 km off the Virginia coast, and two profiles obtained near Bermuda. The average convection current for the five flights is shown by the circles and broken curve in Figure 6. The convection current is obtained from (5), assuming j_e to be constant with altitude and equal to j above 2.5 km. The measured value of j above that altitude is 1.51×10^{-12} amp/m² \pm 3.5 per cent. As shown in Figure 6, the maximum value of j_e obtained experimentally is 1.73×10^{-12} amp/m², and it occurs at 0.015 km, the lowest altitude reached. This is 1.15 times as large as the average total current!

To calculate j_e by means of (7) and (8) it

was necessary to obtain values for $\partial\rho/\partial Z$. This was done by taking the tabulated values of ρ , the potential gradient for these flights and deriving an empirical equation for $\partial V/\partial Z$ which best fitted the average experimental points. This expression was then differentiated twice, using Poisson's equation, and

$$\frac{\partial \rho}{\partial Z} = - \left(1800e^{-20.3Z} + 53.7e^{-1.66Z} + 0.39e^{-0.33Z} \right) \times 10^{-16} \text{ coul/m}^4$$

Values for A in (7) were obtained by using $\partial\rho/\partial Z$ and experimentally determined values of j_e at the same altitude. A straight-line plot of $\log A$ versus $\log Z$ provided values for A_1 and g from (8), resulting in the empirical expression for j_e that is plotted as the solid curve in Figure 6.

Although the experimental values of j_e were used to determine the constants A_1 and g , it may be seen from Figure 6 that the empirical curve does not correspond too well to the experimental curve, particularly at the lower altitudes. Specifically, there is no tendency for the experimental curve to reach a maximum about 0.015 km, although one is indicated in the en-

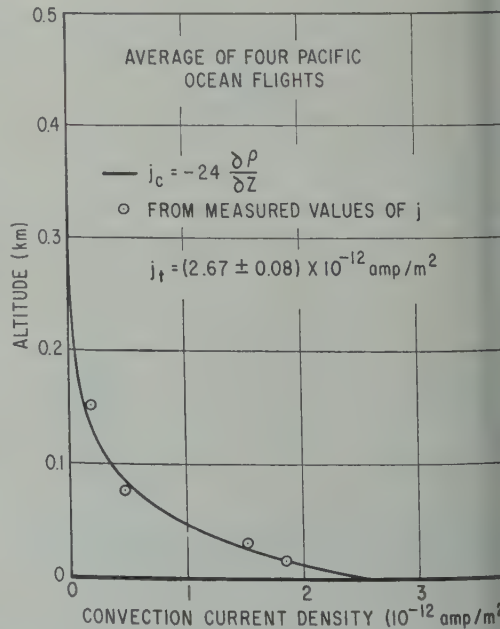


Fig. 7. Average convection current density profile off west coast of the United States.

curve. This may be due to the inadequacy of the theory of turbulence as applied to the problem, or it may be due to the production of space charge near the earth's surface as proposed by Mühleisen [1958].

A similar analysis has been made of four flights obtained off the west coast of the United States. These flights were made 20 km off California in December 1955, between 1900 and 2100 GMT. The average convection current is indicated by the circles in Figure 7. Above 0.3 km j is constant with altitude and has a value of 6.7×10^{-13} amp/m² ± 2.8 per cent. The maximum observed value of j_e , 1.85 units, again occurs at 0.015 km and is 69 per cent of the convection current.

The west coast data were not as amenable to statistical treatment as were those from the Pacific coast, and four exponential terms were required for a reasonable approximation of the potential gradient data. Thus, after differentiating twice,

$$= -\left(1060e^{-20Z} + 4.69e^{-2.3Z} + 0.22e^{-0.9Z} + 0.02e^{-0.23Z}\right) \times 10^{-16} \text{ coul/m}^4 \quad (10)$$

An empirical expression for j_e is given in Figure 9 in which it may be noted that the index of stability g is zero. This may be overly conservative, but it does indicate extreme atmospheric instability, as was observed in the air, and as is supported by the fact that the temperature was not constant up to 2 km.

As may be seen from Figure 7 that the influence of turbulent exchange, with the consequent contribution to convection current, is confined to a thinner stratum off the west coast than off the east coast. This is because of the very low temperature inversion over the Pacific coast. The fit of the empirical curve to the experimental points is better in Figure 7 than in Figure 6. However the zero value for the stability index, its equivalent, the constancy of the exchange coefficient, is not realistic near the surface. This would imply that the maximum value of j_e would occur at the surface, and it is evident that j_e must approach zero at the earth's surface.

However inadequate the application of the

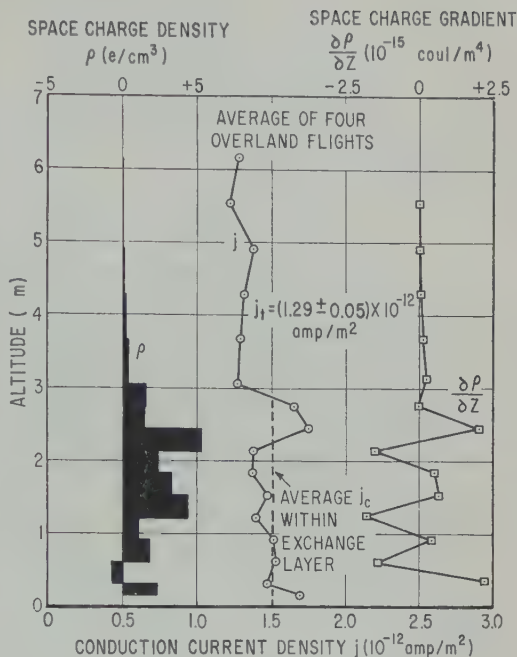


Fig. 8. Space charge convection over land.

theory of turbulent exchange to the convective transport of charge near the earth's surface may be, it is probably the best presently available. Nevertheless the experimental results and the analysis indicate at least two things. First, the convection current density over the ocean is several orders of magnitude greater than that estimated by Gish [1949], down to an altitude of 15 m. Second, there is a strong possibility that surface measurements of current density, normally made by the indirect method at altitudes of 1 m or greater are appreciably influenced by convection.

Four overland soundings were made between March 1955 and May 1956 over Virginia and Maryland plains. The current density profiles from these flights are essentially different from those obtained over the oceans. Plots of j , ρ , and $\partial\rho/\partial Z$ versus altitude are given in Figure 8. It may be seen from the figure that j_e , instead of decreasing with altitude as it does over the ocean, varies above and below an average constant value throughout the surface exchange layer (the average exchange-layer top was about 2.8 km), decreasing rather abruptly at the top. The diffusion of space charge upward from a region rich in positive charge to one deficient

TABLE 1. Summary of Space Charge Convection

	-0.50 to -0.01	0	0.01 to 0.50	0.51 to 1.00	1.01 to 1.50	1.51 to 2.00
j_c/j_t						
No. of occur- rences	2	8	7	9	2	1

in positive charge does not occur over land, and consequently the theory of turbulent exchange cannot be reasonably applied. This may be seen from Figure 8 in which the histogram of ρ seems to be more nearly in agreement with the profile of j_c than does $\partial\rho/\partial Z$. Thus a more constant upward motion of space charge within the exchange layer is more in accordance with the observations than is a decreasing upward diffusive transport of space charge.

These observations could be explained by a rather general upward vertical wind within the exchange layer, but this is difficult to justify meteorologically. A more reasonable explanation might be the attachment of the space charge to atmospheric particles or hygroscopic nuclei which obey a diffusive separation upward, as a result of particle concentration or water vapor gradients, rather than on the weaker space charge gradients. Or it may be due to the upward transport of space charge in columns, in which the contribution to the average conductivity and potential gradient at a constant altitude is small, as suggested by Kraakevik and Clark [1958].

A summary of the effects of convection current for all the flights in this investigation is given in Table 1. This includes seven flights which provided valid data near the earth's surface but were not valid for general analysis. It may be seen from the table that the ratio of maximum convection current to total current density j_c/j_t is negative in only two cases. Thus in but 2 of 29 soundings the upward convection of negative space charge was encountered. Eight soundings showed no trend away from a constant vertical conduction current, as is indicated in column 2. The other columns show that in about two-thirds of the soundings a definite convection current was observed which had an average value of 65 per cent of the total current. For all the flights the average j_c/j_t is 0.40. Since 23 of these soundings were made over

water it is apparent that considerable convection is present over the oceans as well as over the continents.

Because of the complexities of turbulent motion, the application of the theory of turbulent exchange to the transport of physical quantities is made with some trepidation. This is particularly true when it is known that no one mechanism explains all the facts. However, since there is obviously mechanical transport of electric charge, the goals might be reversed. Under controlled, or even normal, fair weather atmospheric conditions, measured values of space charge density might profitably be used to study the turbulence process itself and to determine the exchange coefficients. Electrical characteristics might prove to be a better physical quantity with which to study turbulent transport than momentum, heat, or something else.

5. *Universal and local diurnal effects.* It is the desire of almost every worker in atmospheric electricity to extract from the interrelated mass of variables those quantities which respond to world-wide phenomena and those which are purely local in nature. Since the atmospheric leakage current during fair weather is now generally believed to be continuously replenished by world-wide thunderstorm activity, the problem has become one of adequately and simply monitoring universal thunderstorm activity.

One method of accomplishing this is to determine the atmospheric potential directly, by integrating the potential gradient to the top of the atmosphere, where

$$V_s = - \int_0^\infty E \, dZ \tag{1}$$

and $V_s \rightarrow V$ as $Z \rightarrow \infty$. Clark [1958] has shown that individual measurements of total potential V derived from potential gradient profiles correlate very well with statistical studies of surface potential gradient variations over the oceans. Since Clark and the present author collaborated on measurements, data for current density are available for the same series of flights. It is therefore of interest to see whether current density also provides a useful monitor of universal effects.

The series of five soundings was made over Søndre Strømfjord, Greenland, at various times of day in August and September, 1955. It was hoped that because of the cleanness of the

ere, local effects could be minimized and current density might respond to universal variations. The cleanness of the air is indicated by a low value of average columnar resistance up to 10 km, 6.1×10^{18} ohm-m², where R_s is obtained from the measured conductivity by

$$R_s = \int_0^z \frac{dZ}{\lambda} \quad (12)$$

The values of average current density obtained during each of these flights are compared with the ocean surface potential gradient results from cruise VII of the *Carnegie* [Torreson and Clark, 1946] in Figure 9. The solid curve is obtained from the ocean surface results, and the open circles represent the measured average current density for each flight. Both sets of data have been normalized to their respective diurnal values.

The agreement between the two curves is very good, considering that the solid curve was obtained from many hourly averages and that the open circles represent individual profiles. This indicates that the diurnal variation may be more a statistical effect and may be obtained by averaging observations under proper conditions. However, it must be emphasized that the Greenland current measurements were obtained above the surface exchange layer, small though it was. Furthermore, although the conductivity varied somewhat at a given altitude from flight to flight, the integrated columnar resistance was constant for all flights to within 2 per cent. Consequently the current varied directly with the total potential and presumably with world-wide thunderstorm activity.

A comparison between surface measurements of induction current density and those obtained at altitude was afforded in the same manner as of soundings in Greenland. The surface station, located atop the operations building at the air base (normally upwind of most of the contamination from the base itself), continuously recorded both polar conductivities and potential gradient. The soundings were made at 100 miles from the airfield above the fjord. A comparison between j at the surface and j measured at altitude, together with the potential obtained from Clark [1958], is given in Table 2. The surface station was not started early enough to obtain data during the soundings of August 28. However the current

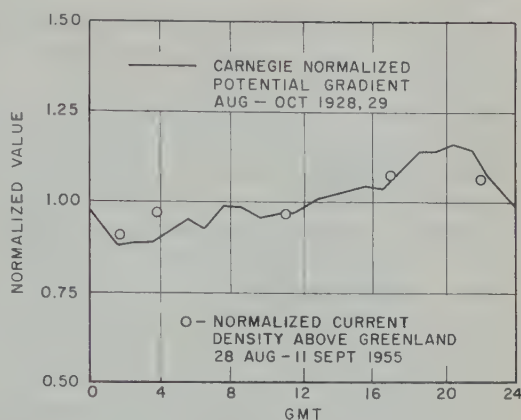


Fig. 9. Diurnal variation in current density over Greenland fjord.

data from two flights conducted on September 9 are included, although no potentials are available because the maximum altitude of the soundings was restricted because of airplane mechanical difficulties and high-altitude cirrus clouds.

It may be seen from Table 2 that the current density aloft follows the total potential variation quite faithfully. However the correlation between the current density results at the surface and aloft is rather poor. It should be noted that the surface values of j are not on an absolute basis because the geometric field augmentation of the building was not known. The variations in surface current density indicate that this quantity cannot be used to determine world-wide effects, except on a statistical basis, i.e., averaging the data from many hourly observations. Perhaps there was some contamination from the airbase; more probably the influence

TABLE 2. Results from Greenland, at the Surface and Aloft

Time GMT	Date	j at Surface (10^{-12} amp/m ²)	j at Altitude (10^{-12} amp/m ²)	V kv
0145	8-31	3.66	3.48	276
0345	9-11	1.75	3.71	275
0630	9-9	4.18	3.26	
1100	9-11	2.80	3.69	283
1215	9-9	1.57	3.13	
1700	8-28		4.12	300
2200	8-30	5.82	4.07	324

TABLE 3. Results of Measurements at the Surface and Aloft near Key West

Quantity	Units	3/27/56 2045 GMT	3/28/56 0600 GMT	3/28/56 2000 GMT
j at surface	10^{-12} amp/m ²	6.87	6.99	6.78
j at altitude	10^{-12} amp/m ²	2.43	2.56	2.60
V	kv	301	216	304
R	10^{16} ohm-m ²	12.7	8.8	11.9
Height of exchange layer	km	0.9	0.6	1.3
Average λ within ex- change layer	10^{-14} mho/m	1.68	3.61	2.16

of convection currents was responsible for the variation in surface values. This indicates that, at a single station, surface measurements of current density and, more emphatically, those of potential gradient and conductivity are not valid as a continuous index of universal effects during fair weather, unless conceivably at even better selected locations.

Another opportunity to compare measurements made at the surface and aloft was afforded during an abbreviated diurnal study over the Gulf of Mexico, 50 km west of Key West, Florida, in March 1956. Three flights were made at the diurnal maximum, minimum, and maximum again, within a 26-hour period. During this time a surface station was recording continuously atop the operations building at Boca Chica Naval Air Station, some 15 km east of Key West.

A comparison of the results is shown in Table 3. The larger surface current density is no doubt due to the field augmentation of the building, which may well be about a factor of 3. No significant changes in j either at the surface or at altitude were observed for the three profiles, and the ratio of the average values, 2.7, may well be the augmentation at the ground site.

In contrast to the Greenland data, the current density at altitude in Table 3 is constant to within 4 per cent for the three flights, whereas the total potential is high and approximately the same at both expected diurnal maxima (about 2000 GMT) and low at the usual minimum. The weighted mean value for V is 259 kv and the variation about this mean is +16, -17, and +17 per cent, respectively. This represents a diurnal variation in agreement with the *Carnegie* results, although the average V is somewhat lower than the average for all 22 soundings reported by *Clark* [1958] as 290 kv.

The reason for the constant j may be seen by examining the total columnar resistance R , where the contribution above 6 km of 2.23×10^{16} units is obtained from the balloon data of *Stell*, *Coroniti*, *Nazarek*, *Kotas*, *Seymour*, and *W* [1955] over New Mexico. The variation in R over Key West is approximately the same as that in V and therefore masks any change in j corresponding to V . This variation in R is a local effect, due primarily to the height of the exchange layer and to the average conductivity within it. This process of greater mixing and higher nucleus densities throughout a thick exchange layer during the afternoon hours has been adequately discussed by *Sagaly* and *Faucher* [1956]. Such a variation in R undoubtedly exists over and near the continent, particularly near population centers. The current density at a given location is therefore generally influenced by such local variations within the entire exchange layer, as well as by the universal variation in total potential. Consequently the use of j as an index of universal effects, particularly in temperate zones near the continent, seems hopeless.

The atmospheric potential results for this investigation, based on equation 11, have already been presented by *Clark* [1958]. However it is of interest here to compare his results with those obtained by a different method to evaluate the assumption of quasi-static equilibrium in the atmosphere and the net effect of turbulent convection on the atmospheric potential. An alternative method for computing involves determining the average current for each sounding, \bar{j} , and multiplying it by the columnar resistance R . Thus

$$V_i = R\bar{j}$$

The value used for \bar{j} for each flight is the av-

TABLE 4. Summary of Atmospheric Potential

Potential	V , kv	σ , %	V_i , kv	σ , %
Actual	263	18.0	248	15.4
Standardized 6 km	301	12.5	299	12.6
Standardized total	245	13.6	232	12.2
	282	9.8	280	11.1

and above the convective region, which is found to be the same up to the high atmosphere.

Results of computing the potential by the methods are given in Table 4 for several soundings. The average potentials at 6 km for fair weather soundings, together with the corresponding standard deviations are given in Table 4. The total potential for V is obtained by eliminating the exponential variation in potential gradient in the 3- to 6-km region to the high atmosphere. This procedure gives a total potential which is probably too low by something less than 10 per cent [Clark, 1958] because it is known that the increase in conductivity and decrease in potential gradient above 6 km is less than indicated [Stergis and others, 1955; Stergis, and Kangas, 1957a]. The increment in columnar resistance above 6 km, obtained from the conductivity data [Woessner, Cobb, and others, 1958; Stergis, and others, 1955], is multiplied by the average current density to provide a potential. This is probably too high because it is the nonvertical current at high altitudes. The total potentials derived in the two ways are compared in line 2 in Table 3. It may be seen that the standard deviation is reduced from line 1, indicating that the potentials tend to converge at high altitudes.

6-km potentials are standardized to their mean values by means of the Carnegie

surface potential gradient results, and the average values are given in line 3. Finally, the average standardized total potentials are given in line 4, where the spread in values is seen to be the least. The table shows that the two methods give results that are in substantial agreement. Thus it may safely be stated that the average conditions found in the fair weather atmosphere provide reasonably valid quasi-static equilibrium above the convective layer. However, Clark's value for average total potential, corrected for nonexponential increase in conductivity with height is $290 \text{ kv} \pm 10$ per cent. The corresponding corrected value for V_i would be $273 \text{ kv} \pm 11$ per cent. The difference of approximately 6 per cent could be due to the average integrated effect of space charge convection.

6. *Geographical studies.* A sufficient number of soundings were made in five different geographical areas to provide some regional electric characteristics of the atmosphere. These results are presented in Table 5, where the diurnal effect is removed from current density and potential to give the standardized \bar{j}_a and V_{ia} . As one reads down the table from the Arctic to eastern United States, the regions are seen to be more and more contaminated. This is evidenced by an increase in total columnar resistance and a corresponding decrease in average current density. The lowest average value of \bar{j} for a single sounding of 1.09×10^{-12} amp/m² was obtained over Maryland. It is of interest to note that the increase in R and the decrease in \bar{j}_a from the Arctic to continental United States is about a factor of 3. The ± 7 per cent variation in V_{ia} indicated in the table is considered to be random. Although R varies diurnally owing to local solar heating, as is shown by the Key West results in the previous section, no attempt has been made to correct for this. Furthermore, although there

TABLE 5. Atmospheric Electric Characteristics of Regions

Region	No. of Flights	Season	GML,* deg	λ_s (10^{-14} mho/m)	R_6 (10^{16} ohm-m ²)	R (10^{16} ohm-m ²)	\bar{j}_a (10^{-12} amp/m ²)	V_{ia} , kv
	6	Su	76	27.1	6.11	7.87	3.85	303
	4	Wi	41	25.5	8.93	11.05	2.55	282
Mexico	3	Sp	36	24.6	8.90	11.13	2.41	269
U. S.	5	Sp	45	26.0	14.03	16.03	1.63	261
	4	Sp-Su	50	26.6	20.21	22.21	1.32	293

*magnetic latitude.

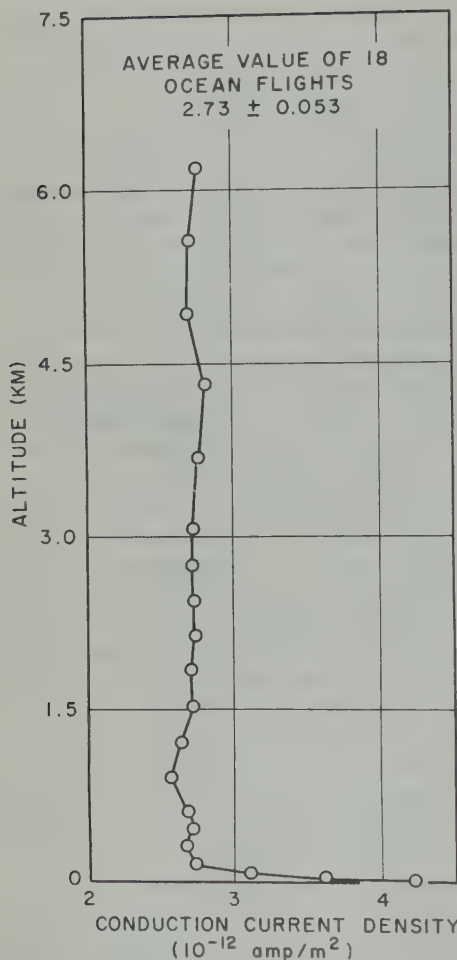


Fig. 10. Average conduction current density profile over the oceans.

is a known seasonal variation in atmospheric potential, the data are insufficient to show anything conclusive about it.

A slight variation in the average conductivity at 6 km λ_s may be seen from Table 5. This may be a real geomagnetic latitude effect, due to the increase in cosmic-ray ionization at higher geomagnetic latitudes (GML). There is a direct correspondence between λ_s and GML, although the effect is not as strong as it might be at even lower latitudes because the knee of the curve of cosmic-ray ionization versus latitude occurs between 40° and 50° . The latitude variation in R is completely masked by the stronger variation due to contamination within the exchange layer.

The average current density distribution for

the 18 ocean soundings is presented in Figure 10. These include the profiles obtained over Søndre Strømfjord, Greenland. All values of j have been standardized to the diurnal mean value.

The average effect of convection on the current density over the ocean is seen in the large value of j at the three lowest altitudes (15 m, 30 m, and 76 m). The convection current density at these altitudes is respectively 54, 33, and 19 per cent of the total current. Excluding the three lowest points, the average value of conduction current density over the ocean is 2.73×10^{-12} amp/m² ± 2.0 per cent. The limits of variation at each altitude are essentially the extremes of the average values of j encountered on the 18 profiles. The average value of j encountered on the Virgin Islands coast to 4.1×10^{-12} amp/m² measured in Greenland. How representative 2.73×10^{-12} amp/m² is of the average j all over the world depends on how well the 18 profiles obtained near North America represent the earth as a whole. This is unknown. However, since the mean value of j is close to the median and approximately midway between the maximum and minimum values, the statistical distribution is good and the average is considered representative.

If the above figure is assumed to be representative of the earth as a whole, the total supply current may be estimated. Although lower over populated land areas, it is higher over mountainous areas where the atmospheric columnar resistance is lower and the current is augmented by the mountain peaks themselves. Hence the ocean mean value for j may be fairly close to the average for the whole earth. If the average j is multiplied by the earth's area, neglecting the area of thunderstorm activity, the total world-wide fair weather current is 1800 amp. This estimate is appreciably lower than Gish's [1951] estimate of 1800 amp. Since this estimate was based on surface measurement of conduction current density, it is quite possible that his average value is some 30 per cent lower owing to convection. If the 1400-amp value is correct, and it be assumed that the number of active thunderstorms which generate the supply current is 1800, as estimated by Brooks [1957], then the average current per thunderstorm to maintain equilibrium is 0.8 amp. This is in better agreement with the average current per thunderstorm measured by Stergis, Rein, and Katsenberger [1957b] above Florida thunderstorms.

clusions. An investigation of the conduction current density in the atmosphere on 22 flights made from aircraft in the altitude range from 15 m to 6 km shows:

Above the exchange layer the conduction current density on a given sounding is constant with altitude to within about 10 per cent.

Within the exchange layer a convection current exists, and over the oceans it is produced by the upward diffusion of positive space charge. It generally decreases with altitude, having its maximum value at or below 100 m. For some soundings it has a magnitude of up to twice the total current density. Conduction currents up to 50 per cent of the total current sometimes occur at the top of the exchange layer, when the space charge density and electric field are high and turbulence is present. Five soundings off the Atlantic coast and four off the Pacific coast have been analyzed in the light of the theory of turbulent exchange. Although turbulent diffusion is the most likely explanation of the observed data, it does not completely account for the high convection current density at the lowest altitudes reached (15 m), which off the east coast averages 115 per cent and off the west coast 69 per cent of the total current density. However, it is suggested that the process might be reversed and that space charge density might be used as a tool to study the turbulent exchange process.

Within the exchange layer over land, the convection current on four soundings averaged randomly about a mean value of 20 per cent of the total current. Upward diffusion of positive space charge cannot account for the observed data. One possible explanation is the attachment of space charge to particles of condensation nuclei, which obey their own laws of diffusive separation.

Over Greenland the total atmospheric potential and the average current density aloft vary with time, and both are in phase with the universal diurnal variation in thunderstorm activity. However, surface measurements of current density show a poor correlation with the potential. Near Key West the columnar current density varies diurnally in opposite phase to the variation in potential, and so the current density remains sensibly constant with time. Correlation with surface measurements is again

poor. It is concluded that, although current density measured at the earth's surface may show universal effects if data are averaged over many days, it cannot be used as a day-to-day index of world-wide variations.

6. The average current density in arctic regions is 3.9×10^{-13} amp/m²; over the eastern United States it is 1.3×10^{-12} amp/m². This range brackets average values measured in other areas. The decrease by a factor of 3 corresponds inversely to an increase in total columnar resistance from 7.9 to 22.2×10^{16} ohm-m², indicating the range of pollution from clean arctic to contaminated continental air mass. The conductivity at 6 km increases with geomagnetic latitude, suggesting that it is caused by the increase in cosmic-ray ionization with latitude.

7. The average current density distribution over the oceans indicates a convection current of as much as 54 per cent of the total current near the earth's surface. However, it has an average value of 2.73×10^{-12} amp/m² above 100 m, constant to within 2 per cent. If this value is representative of the earth as a whole, the total world-wide conduction current is 1400 amp from the high atmosphere to the earth.

Acknowledgements. It is a pleasure to thank Dr. John F. Clark for his cooperation in the experimental program and for the use of his potential gradient results. Thanks are also due Messrs. G. P. Serbu and R. V. Anderson, and pilots CDR R. L. Milner and LCDR M. J. Randleman, who contributed substantially in obtaining the flight data, and to Mr. E. C. Whipple, Jr., and Mrs. Eva Mae Trent, who assisted in the analysis.

REFERENCES

- Brooks, C. E. P., The distribution of thunderstorms over the globe, *Geophys. Mem.*, 3, 145-164, 1925.
- Clark, J. F., The fair-weather atmospheric potential and its gradient, in *Recent Advances in Atmospheric Electricity*, edited by L. G. Smith, Pergamon Press, London, pp. 61-73, 1958.
- Everling, E., and A. Wigand, Spannungsgefälle und vertikaler Leitungsstrom in der freien Atmosphäre, nach Messungen bei Hochfahrten im Freiballon, *Ann. Phys.*, 66, 261-282, 1921.
- Gish, O. H., Atmospheric electricity, in *Terrestrial Magnetism and Electricity*, edited by J. A. Fleming, Dover Publications, New York, pp. 149-230, 1949.
- Gish, O. H., Universal aspects of atmospheric electricity, in *Compendium of Meteorology*, edited by T. F. Malone, American Meteorological Society, Boston, pp. 101-119, 1951.

- Gish, O. H., and G. R. Wait, Thunderstorms and the earth's general electrification, *J. Geophys. Research*, **55**, 473-484, 1950.
- Holzer, R. E., Studies of the universal aspects of atmospheric electricity, 1951-1955, *Final Rept., contract AF19(122) -254*, University of California, Los Angeles, p. 28, 1955.
- Kasemir, H. W., Zur Strömungstheorie des luftelektrischen Feldes III: Der Austausch-generator, in *Atmospheric Electricity during IGY* (Report of meeting at Aachen, May 11-12, 1956), p. 13, 1957.
- Kraakevik, J. H., The airborne measurement of atmospheric conductivity, *J. Geophys. Research*, **63**, 161-169, 1958a.
- Kraakevik, J. H., Electrical conduction and convection currents in the troposphere, in *Recent Advances in Atmospheric Electricity*, edited by L. G. Smith, Pergamon Press, London, pp. 75-88, 1958b.
- Kraakevik, J. H., and J. F. Clark, Airborne measurements of atmospheric electricity, *Trans. Am. Geophys. Union*, **39**, 827-834, 1958.
- Mühleisen, R., The influence of water on the atmospheric electric field, in *Recent Advances in Atmospheric Electricity*, edited by L. G. Smith, Pergamon Press, London, pp. 213-221, 1958.
- Sagalyn, R. C., and G. A. Faucher, Space and time variations of charged nuclei and electrical conductivity of the atmosphere, *Quart. J. Meteorol. Soc.*, **82**, 428-445, 1956.
- Schonland, B. F. J., *Atmospheric Electricity*, Wiley & Sons, New York, p. 29, 1953.
- Stergis, C. G., S. C. Coroniti, A. Nazarek, I. Kotas, D. W. Seymour, and J. V. Werme, Conductivity measurements in the stratosphere, *Atmospheric Terrest. Phys.*, **6**, 233-242, 1954.
- Stergis, C. G., G. C. Rein, and T. Kangas, Electric field measurements in the stratosphere, *J. Atmospheric and Terrest. Phys.*, **11**, 77-82, 1955.
- Stergis, C. G., G. C. Rein, and T. Kangas, Electric field measurements above thunderstorms, *J. Atmospheric and Terrest. Phys.*, **11**, 83-90, 1955.
- Sutton, O. G., *Micrometeorology*, McGraw-Hill Book Co., New York, p. 56, 1953.
- Torreson, O. W., and others, *Scientific Results of Cruise VII of the Carnegie during 1928-29*, Carnegie Inst. Wash. Publ. 568, p. 163, 1946.
- Tverskoi, P. N., and M. P. Timofeev, Turbulence and the vertical profile of electric field intensity in the lower layer of the atmosphere, *Izv. Akad. Nauk SSSR, Ser. Geograf. i Geofiz.*, **12**, 377-386, 1948.
- Woessner, R. H., W. E. Cobb, and R. Gunn, Simultaneous measurements of the positive and negative light-ion conductivities to 26 kilometers, *Geophys. Research*, **63**, 171-180, 1958.

(Manuscript received June 26, 1961.)

Atmospheric Emission and Opacity at Millimeter Wavelengths Due to Oxygen

M. L. MEEKS

Georgia Institute of Technology, Atlanta, Georgia

Abstract. The theory of Van Vleck and Weisskopf has been used to compute the emission and opacity of the earth's atmosphere in the wavelength region from 7.5 to 3.7 mm. Equal pressure broadening was assumed for the 25 oxygen transitions which occur in this spectral region. The line-broadening parameter $\Delta\nu$ was set equal to 600 Mc/s for $T = 300^\circ\text{K}$ and $p = 1$ atmosphere, and the temperature dependence was assumed to be proportional to $T^{-0.65}$. Digital computer calculations were made of τ_0 , the total zenith opacity of the atmosphere as a function of frequency ν . In these computations the ARDC model atmosphere was used, and excellent agreement was found with measured values of τ_0 at 6-mm and 4.3-mm wavelengths. Emission was computed similarly in terms of $T_0(\theta)$, the antenna temperature contributed by oxygen for observation at a zenith angle θ . Computations were made of τ_0 and T_0 for elevations of 0, 2.5, 3.5, 8, and 30 km above the earth in order to determine the advantages of various sites for astronomical observations in this wavelength region. The radiation from the atmosphere to space was also computed, and the oxygen lines were found to appear in emission.

Introduction. Propagation through the atmosphere at millimeter wavelengths is easily controlled by the absorption of oxygen and water vapor. Oxygen at temperatures characteristic of the atmosphere has 25 absorption lines around the wavelength of 5 mm and an additional line near 2.5 mm. At low levels in the atmosphere, pressure broadening blends the complex of lines at 5 mm together to form a continuous region of absorption. On the other hand, water vapor has only two lines in the millimeter region, one at 13.39 mm and another at 6 mm. However, there are a large number of additional H_2O lines at submillimeter wavelengths, and the skirts of these pressure-broadened lines produce continuous absorption at the short-wavelength end of the millimeter

region of these molecules in air. An additional difficulty associated with such computations stems simply from the large volume of numerical operations which are necessary to handle such an analysis. However, large-scale digital computers now make it possible to perform such computations and to correlate measurements in the laboratory at low pressures with measurements in the atmosphere.

Of the two important atmospheric constituents, O_2 and H_2O , there appears to be more complete information on the pressure-broadening properties of O_2 , particularly in the case of the 25 O_2 lines around 5 mm. For this reason the computations reported here have been restricted to oxygen in the region between 50 and 80 gc/s,¹ which is dominated by O_2 absorption. Somewhat similar computations for oxygen and water vapor have been reported by Hogg [1959] for the frequency range 0.5 to 40 gc/s.

2. The oxygen spectrum. The millimeter spectrum of O_2 consists of a series of transitions between fine-structure levels of various rotational states of the molecule.² Oxygen in its ground

state has been a steadily increasing interest in the propagation of millimeter waves through the atmosphere in recent years, both for radio astronomy and engineering applications. So it is of interest to compute the emission and absorption properties of the atmosphere based on a knowledge of the spectra of O_2 and H_2O in the millimeter region. There are, however, certain difficulties in carrying out such computations, although the structure and spectrum of O_2 and H_2O are well understood. These difficulties are primarily associated with the pressure broad-

¹ The frequency unit of 1 gigacycle/sec (gc/s) = 10^9 cps is used here in place of the somewhat more familiar but longer equivalent, the kilomegacycle/sec.

² For a complete discussion of the O_2 spectrum, see Townes and Schawlow [1955].

electronic state is coupled to the electromagnetic field by a magnetic moment, which results from the unpaired spins of two electrons. The electron-spin quantum number is $S = 1$, and the molecular state is $^3\Sigma$. Only states with odd rotational quantum number $N = 1, 3, 5, 7, \dots$ are permitted, and the transitions of interest are of two types, $J = N \rightarrow J = N + 1$ and $J = N \rightarrow J = N - 1$, where J is the total angular momentum quantum number. The frequencies of these transitions will be referred to respectively as ν_{N+} and ν_{N-} . These frequencies have been determined with precision by *Mizushima and Hill* [1954], and the experimental spectrum has been used by *Tinkham and Strandberg* [1955] to obtain a best fit to five parameters of the O_2 molecule. Table 1 shows the transition frequencies computed by Tinkham and Strandberg after the evaluation of the molecular parameters.

In computing the absorption coefficient γ for oxygen the Van Vleck-Weisskopf line shape is used [*Van Vleck and Weisskopf*, 1946]. The required calculation has been specified in detail by *Van Vleck* [1946], and the final expression for γ in air as a function of frequency ν (in gc/s), P (in mm of Hg), and temperature T (in degrees K) is given below.

$$\gamma(\nu, P, T) = C_1 P T^{-3} \nu^2$$

$$\cdot \sum_N \{F_{N+\mu_{N+}}^2 + F_{N-\mu_{N-}}^2 + F_{0\mu_{N0}}^2\} A_N \quad (1)$$

where

$$C_1 = 2.6742 \text{ for } \gamma \text{ (in db/km).}$$

$$C_1 = 0.61576 \text{ for } \gamma \text{ (in nepers/km).}$$

$$F_{N\pm} = \frac{\Delta\nu}{(\nu_{N\pm} - \nu)^2 + \Delta\nu^2} + \frac{\Delta\nu}{(\nu_{N\pm} + \nu)^2 + \Delta\nu^2}.$$

$$F_0 = \frac{\Delta\nu}{\nu^2 + \Delta\nu^2}.$$

$$\mu_{N\pm}^2 = \frac{N(2N+3)}{N+1},$$

$$\mu_{N-}^2 = \frac{(N+1)(2N-1)}{N}.$$

$$\mu_{N0}^2 = \frac{2(N^2 + N + 1)(2N - 1)}{N(N + 1)}.$$

$$A_N = \exp(-C_2 N(N+1)/T), \quad C_2 = 2.0684.$$

TABLE 1. Transition Frequencies for Oxygen

N^*	ν_{N+} , Mc/s	ν_{N-} , Mc/s
1	56 264.7	118 750.7
3	58 446.9	62 486.7
5	59 591.5	60 306.1
7	60 435.5	59 164.0
9	61 151.3	58 323.6
11	61 800.9	57 612.1
13	62 411.9	56 967.8
15	62 998.5	56 363.1
17	63 568.7	55 783.6
19	64 127.6	55 221.5
21	64 678.2	54 671.6
23	65 222.7	54 130.9
25	65 762.6	53 597.3

The line-breadth parameter $\Delta\nu$ is the half-width of a low-pressure line measured to half-maximum. It may be noted that the line shape is a result of resonant and antiresonant terms in $F_{N\pm}$, as well as the nonresonant term involving F_0 . The μ^2 factors are essentially matrix elements, and A_N represents the Boltzmann weight factor.

The line-broadening parameter $\Delta\nu$ for O_2 has been determined to be nearly independent of rotational quantum number N and is assumed to be given by

$$\Delta\nu = \alpha P [0.21 + 0.78\beta] (300/T)^{0.85}$$

For pure O_2 the factor in brackets above would be equal to 1. However, for air this factor consists of the mole fraction of O_2 in air, 0.21, plus the mole fraction of N_2 in air times β , which specifies the relative effectiveness of N_2 collisions to O_2 - O_2 collisions in pressure broadening. The parameters α and β must be determined empirically. The $T^{-0.85}$ dependence of line broadening on temperature is also empirical and is based on measurements of *Hill and Gordy* [1954].

The parameter α in (2) represents the quantity $\Delta\nu/P$ for pure oxygen at a temperature $T = 300^\circ\text{K}$. This parameter has been measured independently at low pressures by *Artman and Gordon* [1954] and by *Hill and Gordy* [1954]. These measurements are in good agreement and yield an average value of $\alpha = 1.95 \text{ Mc/s (mm of Hg)}^{-1}$. Variations in $\Delta\nu$ from line to line were of the order of experimental error. For pressures approaching 1 atmosphere the individual lines blend together around the wavelength of 5 μ .

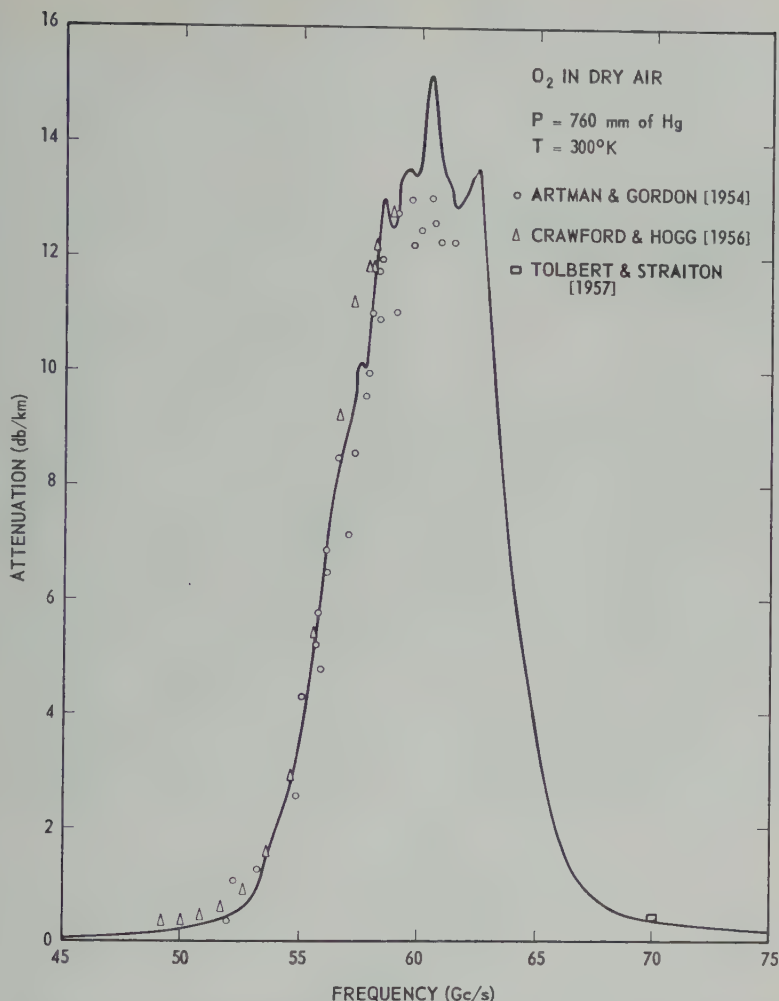


Fig. 1. The attenuation of dry air at sea level. The curve was computed for $\beta = 0.25$, which gives $\Delta\nu = 600$ Mc/s for 1 atmosphere pressure at 300°K. Data from cavity measurements of Artman and Gordon [1954] and free-space attenuation measurements of Crawford and Hogg [1956] and of Tolbert and Straiton [1957] are shown for comparison.

it becomes impossible to measure the widening of a single line.

Several independent measurements have been made of the absorption coefficient of air at 1 atmosphere pressure. These measurements are in good agreement with each other and have been of two types: cavity measurements by Artman and Gordon [1954] and free-space measurements by Crawford and Hogg [1956] and by Tolbert and Straiton [1957]. The value of $\Delta\nu/P$ which must be assumed in order to obtain agreement between measured and computed values of γ , has been estimated to lie roughly between 0.8 and 1.0 Mc/s (mm of Hg) $^{-1}$ [Craw-

ford and Hogg, 1956]. This is roughly half the line broadening to be expected from an atmosphere of pure oxygen. However, by a suitable choice of β in (2), these results can be brought into agreement. The appropriate value of β lies between 0.21 and 0.37, considerably smaller than the value $\beta = 0.90 \pm 15$ per cent, which was estimated from low-pressure measurements of Artman and Gordon [1954]. Nevertheless the selection of $\alpha = 1.95$ Mc/s (mm of Hg) $^{-1}$ in (2), together with the appropriate small value of β , appears to be the most nearly consistent procedure, and the assumption has been made in performing the computations reported here. The

TABLE 2. Assumed Elevations and Sites for Radiometric Observations

Elevation, km	Description of Site
0	Sea level
2.5	High mountain (e.g., Sacramento Peak, New Mexico)
3.5	Very high mountain (e.g., Jungfrauoch, Switzerland)
8.0	Propeller-driven aircraft
30.0	Balloon

detailed examination of the emission and opacity of the atmosphere from 40 to 80 gc/s. This examination has been conducted primarily with view toward the development of radio astronomy in this region of the spectrum.

A decrease in the atmospheric opacity at millimeter wavelengths can be obtained by placing the radiometric apparatus at elevations well above sea level. This is particularly important near the wavelength of 5 mm where the opacity makes sea level observations impossible. Computations have therefore been made for five different elevations. These elevations are listed in Table 2 along with descriptions of the sites. The reduction in opacity with increasing height of the observing site has not been evaluated for water vapor, but total absorption should fall off much more rapidly with height for this constituent of the atmosphere.

The ARDC model atmosphere [U. S. Air Force, 1960] was used to obtain a representative distribution for temperature and pressure in the earth's atmosphere. Figure 2 shows the distributions with selected values of the line broadening parameter $\Delta\nu$ indicated along the pressure curve. At heights above 55 km, $\Delta\nu$ is well below 1 Mc/s, and absorption is negligible.

value $\beta = 0.25$ has been assumed because this value gives the best fit to measured values of the total zenith attenuation (see section 3). The choice of $\beta = 0.25$ corresponds to $\Delta\nu = 600$ Mc/s for $P = 1$ atmosphere and $T = 300^\circ\text{K}$. Figure 1 shows the absorption coefficient γ , which was computed as a function of frequency for air at a temperature of 300°K and pressure of 760 mm of Hg. Measured values of γ are also plotted on this figure, and the agreement may be seen to be generally good.

3. Emission and opacity of the atmosphere. The ability to make rapid machine computations based on equations 1 and 2 makes possible a

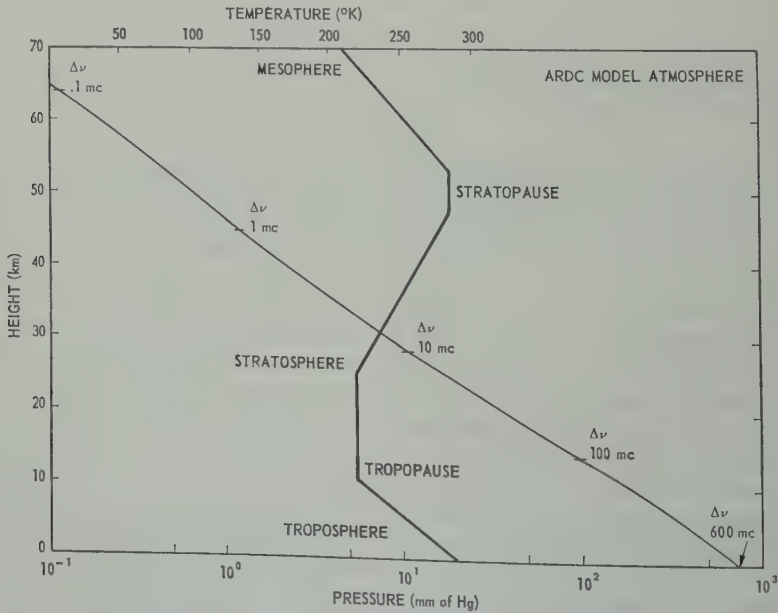


Fig. 2. The temperature and pressure distributions in the ARDC model atmosphere. At several points along the pressure curve representative values of $\Delta\nu$ are given for the indicated heights and pressures.

lines. Integrations have not been carried out at this height.

The opacity τ of the atmosphere is a function of height h above sea level and the zenith angle θ of the line of observation. The opacity $\tau_0(h)$ is given by

$$\tau = \int_h^{\infty} \gamma(r) \sec \theta \, dr = \tau_0(h) \sec \theta \quad (3)$$

where $\tau_0(h)$ is the optical depth measured vertically through the atmosphere to an elevation h above sea level. Figure 3 shows the computed values of τ_0 as a function of frequency for the selected elevations. The ranges of measured values of τ_0 at 50 gc/s by *Whitehurst, Copeland, and Mitchell* [1957] and at 70 gc/s by *Coates* [1958] are shown on the two lower plots. As mentioned in section 2, the value

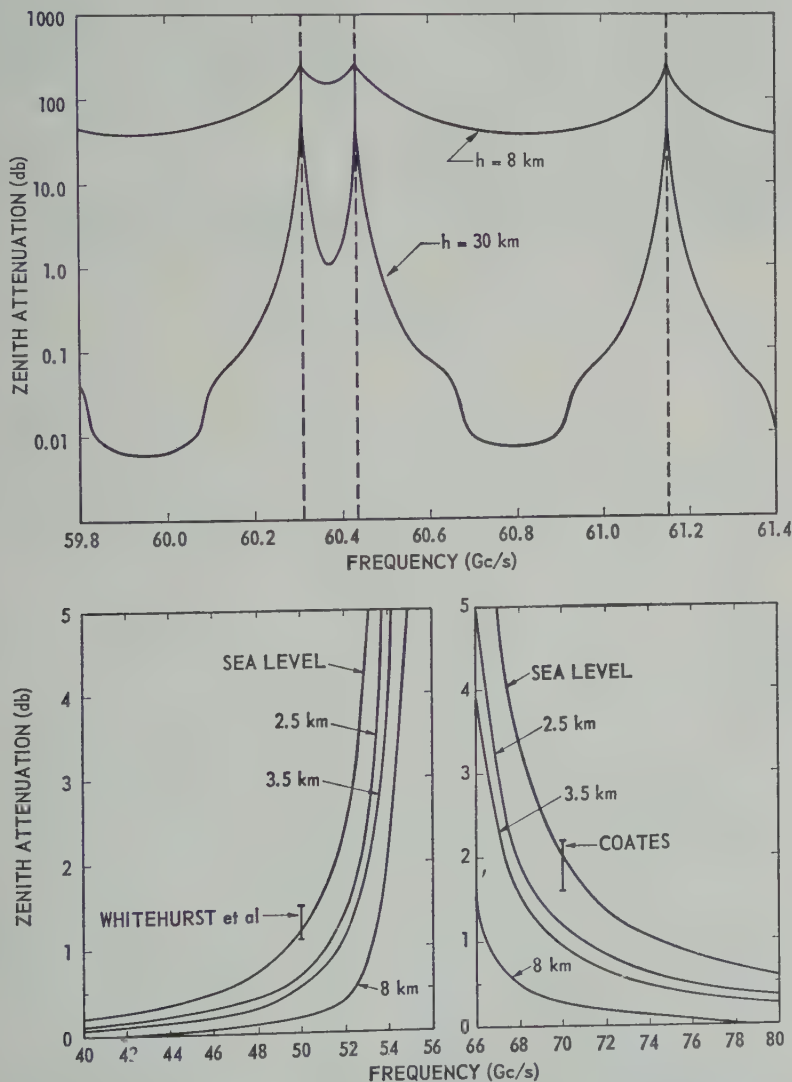


Fig. 3. Computed total vertical opacity due to O_2 for selected heights and frequency intervals. The lower left-hand curves show total vertical opacity versus frequency for the low-frequency neighborhood of the oxygen lines for four elevations. The lower right-hand curves show corresponding results for the high-frequency neighborhood. The upper curves show a sample of the opacity variations at high elevations computed in the vicinity of three oxygen lines near the center of the 5-mm spectrum. Ranges of two sets of measurements are also shown.

$\beta = 0.25$ was selected because it gave the best fit to these data. However, it is encouraging that both these measurements appear to be mutually consistent and the value of β lies in the range 0.21 to 0.37 which was estimated on the basis of measurements of γ at 1 atmosphere. Figure 3 also shows that on high mountains the opacity is about one-half $\tau_o(0)$ over the ranges 40 to 53 gc/s and 68 to 80 gc/s. Observations from propeller-driven aircraft give even greater reductions. The upper part of Figure 3 shows a small portion of the spectrum in the center of the oxygen absorption region, and it may be seen that it would be possible to make radiometric observations through the atmosphere from altitudes around 30 km at frequencies which lie between the oxygen lines. Propeller-driven aircraft at about 8 km are not suitable platforms for such observations because the

opacity is above 30 db even in the region between the oxygen lines.

In addition to the opacity, it is also necessary to determine the emission of radiation from O_2 . For radiometric purposes it is simplest to consider this emission as it contributes to the equivalent radiation temperature T which would be measured with a narrow-beam antenna 'looking' through the atmosphere. If the average radiation temperature of a source beyond the atmosphere is T_s and if zenith angle of the source is θ , the effect of the atmosphere may be described in terms of the zenith opacity τ_o and the radiation temperature $T_o(\theta)$ as

$$T = T_s e^{-\tau_o \sec \theta} + T_o(\theta) \quad (1)$$

The contribution $T_o(\theta)$ to the antenna temperature may be obtained in terms of the atmospheric temperature distribution by means of

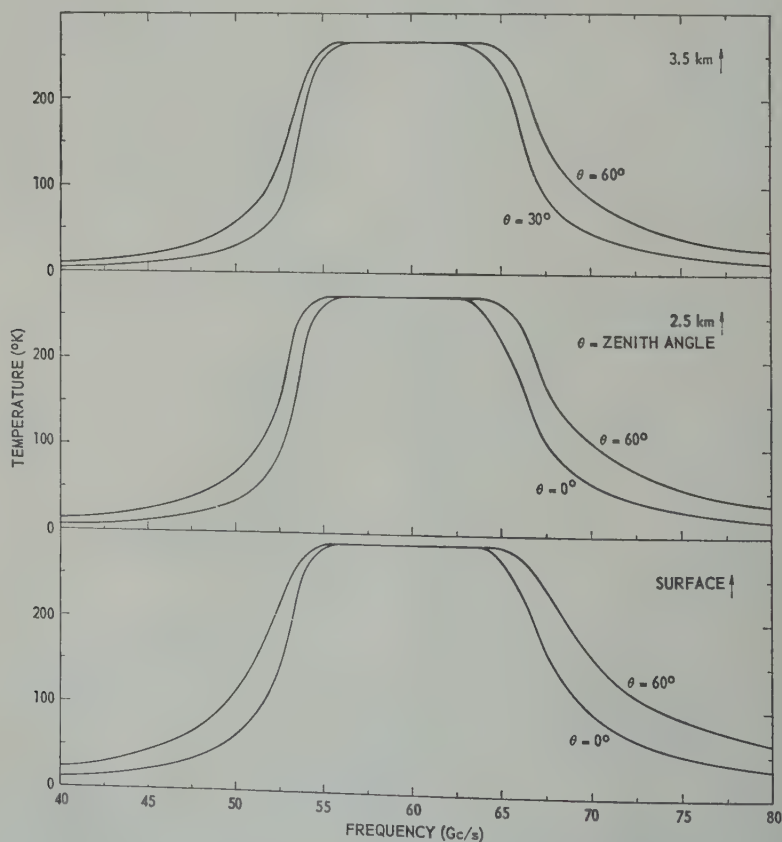


Fig. 4. Computed antenna temperatures versus frequency for antennas at three elevations directed vertically upward and directed at a zenith angle of 60° . These antenna temperatures are due to oxygen alone.

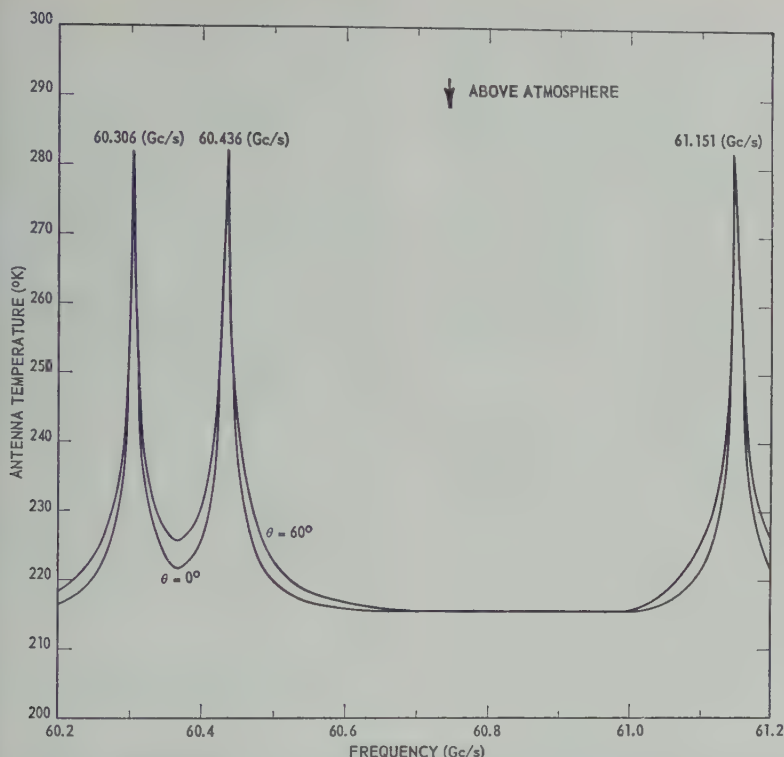


Fig. 5. Computed antenna temperatures for observation of the earth's atmosphere from altitudes above 50 km. The frequency range includes three oxygen absorption lines. Curves are shown for antennas directed vertically downward ($\theta = 0^\circ$) and directed obliquely downward ($\theta = 60^\circ$).

$$T_o(\theta) = \int_0^\infty T e^{-\tau} \gamma \sec \theta \, dh \quad (5)$$

The integral in this equation was evaluated numerically for oxygen in the ARDC model atmosphere. Figure 4 shows the results of this calculation for three observing elevations: sea level, 1.5 km, and 3.5 km. The curves represent zenith angles, $\theta = 0^\circ$ and $\theta = 60^\circ$. In the neighborhood of the O_2 transition frequencies, the absorption is high, the antenna temperature becomes independent of zenith angle and equal to the ambient air temperature. Away from this region of high opacity, the antenna temperature $T_o(\theta)$ increases with increasing zenith angle θ . The dependence of T_o on θ as found from (5) does not agree with the simple case of an isothermal atmosphere where $T_o = T_a = \text{const}$, namely

$$T_{o,i}(\theta) = T_a(1 - e^{-\tau_a \sec \theta}) \quad (6)$$

where, an expression of this form can be

used to estimate $T_o(\theta)$ to within about 2 per cent for zenith angles less than 60° . The value $T_a = 275^\circ\text{K}$ should be used rather than the surface air temperature which is 288°K for the ARDC model atmosphere.

The antenna temperature due to atmospheric oxygen may also be computed for satellite or rocket observations looking down into the atmosphere from above. Equation 5 applies to this case, provided the integration through the atmosphere proceeds in the proper direction. For the computer program used in these calculations, it was necessary only to reverse the order in which data representing the model atmosphere were fed into the computer. Figure 5 shows the computed antenna temperature looking down into the model atmosphere from above 55 km. The frequency region 60.2 to 61.2 gc/s was chosen to include three typical oxygen transitions near the center of the oxygen absorption region. The resonant frequencies in Figure 5 show up in emission with maxima correspond-

ing to the temperature maximum of 282°K at heights around 50 km in the model atmosphere (Fig. 2). Off the resonant frequencies, radiation comes from lower down in the atmosphere, mainly from the isothermal region at 217°K lying between heights of 11 and 25 km.

Two angles of observation are included in Figure 5; these represent observation vertically downward ($\theta = 0^\circ$) and obliquely downward at an angle $\theta = 60^\circ$ with the vertical. The similarity of the temperature versus frequency curves for these two directions indicates no strong dependence on angle θ .

4. *Summary and comments.* The computations of emission and opacity for atmospheric O_2 are consistent with measured values of the vertical opacity at 50 and 70 gc/s and are also consistent with measurements of the absorption coefficient of air at pressures of 1 atmosphere. However, some inconsistency remains between measurements of O_2 absorption at low pressures (with various amounts of N_2 added) and measurements of air absorption at 1 atmosphere pressure. This inconsistency involves the question of oxygen line broadening by oxygen-oxygen and nitrogen-oxygen collisions. Additional laboratory studies of this problem appear to be needed.

Investigation of the advantages of high-altitude observations for millimeter-wavelength astronomy show that altitudes of around 30 km or greater will be necessary to make observations toward the center of the oxygen absorption region near 60 gc/s. Outside of this region of strong absorption, observations from high mountains (altitude 2.5 to 3.5 km) would give about half the sea level opacity of O_2 . Observations from propeller-driven aircraft at 8 km would give much smaller opacities, but the atmosphere remains effectively opaque ($\tau_v > 30$ db) even from this altitude at frequencies near 60 gc/s.

When the earth's atmosphere is observed from a rocket or satellite, the oxygen lines will appear in emission. At the center of a typical line near 60 gc/s, the radiation temperature will be about 280°K as compared with temperatures of about 220°K between the lines (see Fig. 5). This result follows from the thermal structure of the atmosphere, which shows a temperature maximum of about 280°K near 50 km due to ultraviolet ozone absorption and an underlying isothermal region at about 220°K.

The computed examples that show the antenna temperature due to O_2 as a function of frequency and zenith angle (Figs. 4 and 5) contain considerable information about the atmosphere. If precision measurements of T_v could be made as a function of zenith angle at several frequencies, these data could be analyzed to obtain information on the thermal structure of the atmosphere. A proposal of this nature involving infrared measurements has been made by Kaplan [1959]. He proposes use of CO_2 and H_2O around the 15- μ wavelength. Insofar as oxygen is concerned, however, it will be necessary to investigate pressure broadening in more detail before the full potentialities of these measurements can be achieved.

It is clearly desirable to make an analysis of the emission and opacity of water vapor along the same lines that have been followed here for oxygen. An analysis of this kind is planned for the near future.

Acknowledgments. This work was principally supported by a National Science Foundation grant (NSF-G15063) to Georgia Tech for research in submillimeter astronomy. I wish to acknowledge the advice and support I have received from A. E. Lilley, initially and throughout this search. Important contributions to this work were also made by Mr. Frank N. Bash. All machine computations were performed on the Burroughs 220 at the Rich Electronic Computer Center, Georgia Institute of Technology. The computations required 14 hours of computer time.

REFERENCES

- Artman, J. O., and J. P. Gordon, Absorption of microwaves by oxygen in the millimeter wavelength region, *Phys. Rev.*, **98**, 1237-1245, 1955.
- Coates, R. J., Measurements of solar radiation and atmospheric attenuation at 4.3-millimeter wavelength, *Proc. IRE*, **46**, 122-126, 1958.
- Crawford, A. B., and D. C. Hogg, Measurement of atmospheric attenuation at millimeter wavelengths, *Bell System Tech. J.*, **35**, 907-916, 1960.
- Hill, R. M., and W. Gordy, Zeeman effect and line breadth studies of the microwave lines of oxygen, *Phys. Rev.*, **93**, 1019-1022, 1954.
- Hogg, D. C., Effective antenna temperature due to oxygen and water vapor in the atmosphere, *J. Appl. Phys.*, **30**, 1417-1419, 1959.
- Kaplan, L. D., Inference of atmospheric structure from remote radiation measurements, *J. Soc. Am.*, **49**, 1004-1007, 1959.
- Mizushima, M., and R. M. Hill, Microwave spectrum of O_2 , *Phys. Rev.*, **93**, 745-748, 1954.
- Tinkham, M., and M. W. P. Strandberg, Temperature of the fine structure of the molecular oxygen ground state, *Phys. Rev.*, **97**, 937-951, 1955.

C. W., and A. W. Straiton, Experimental measurement of the absorption of millimeter waves over extended ranges, *IRE Trans. Antennas and Propagation*, AP-5, 239-241,

C. H., and A. L. Schawlow, *Microwave spectroscopy*, McGraw-Hill Book Co., New York, 1955.

Force, *Handbook of Geophysics*, revised Macmillan Co., New York, pp. 1-10 through 1960.

Van Vleck, J. H., The absorption of microwaves by oxygen, *Phys. Rev.*, 71, 413-424, 1946.

Van Vleck, J. H., and V. F. Weisskopf, On the shape of collision-broadened lines, *Revs. Modern Phys.*, 17, 227-236, 1946.

Whitehurst, R. N., J. Copeland, and F. H. Mitchell, Solar radiation and atmospheric attenuation at 6-mm wavelength, *J. Appl. Phys.*, 28, 295-298, 1957.

(Manuscript received July 29, 1961.)

Chemical Composition of Rain Water in Kampala, Uganda, and Its Relation to Meteorological and Topographical Conditions

SIMON VISSER

Makerere College, Kampala, Uganda

Abstract. The amounts of Na^+ , K^+ , Ca^{2+} , Fe^{3+} , Mn^{2+} , NH_4^+ , total N, NO_3^- , Cl^- , SO_4^{2-} , and PO_4^{3-} , in daily rain-water samples collected during the course of a year were determined. The sources of these ions in the water are discussed, and it is concluded that local dust and Lake Victoria are the principal contributors. Possible correlations of chemical composition with amount of precipitation, wind direction, and lightning flashes were assessed.

Introduction. There is little information available concerning the chemical composition of local rain water. It was therefore of interest to determine the various cations and anions present in the rain that falls at Kampala, Uganda. Kampala ($0^\circ 25' \text{N}$, $32^\circ 35' \text{E}$) is situated just north of the equator at an altitude of about 1250 m. The average annual rainfall is 1260 mm. There are two very weakly developed seasons, one from March to May (with an average monthly rainfall of 100 to 200 mm) and another from September to November (average monthly rainfall of 100 to 130 mm). During the wet periods the average rainfall varies between 100 and 150 mm per month. There are generally heavy showers and thunderstorms.

The area is influenced by Lake Victoria with a surface area of 69,484 km^2 from which it is only 10 km distant. The wind direction varies considerably during the day because of the lake. Rain water was collected daily from August 1958, through July 31, 1959.

Method of examination. Rain water was collected in a P.V.C. coated metal funnel with a collecting area of 1 m^2 to which a 2.5-liter glass container was attached. The funnel was placed on top of a building about 10 m above ground level.

During the day the samples were collected immediately after each shower; rain that fell at night was collected at 8.30 A.M. the next morning. Week-end rain was collected each Monday morning. Rainfall and wind data were recorded by the Entebbe Meteorological Service. All figures are from Namirembe Hill, about 1 km from the sampling site. In view of the

strongly varying wind directions at lower levels, only the more constant wind directions at 1600 to 2500 m above the land surface were considered. Information about lightning flashes was provided by the Uganda Electricity Board. The flashes were registered by an E.R.A. pattern lightning flash counter. The numbers recorded are the flashes to earth per day within a 40-km radius of Kampala.

Before the chemical analysis was performed the samples were filtered through a Whatman no. 541 paper. The following methods of analysis were used: pH was measured with a Cambridge or Marconi meter; conductivity was measured with a Dionic water tester; potassium, sodium, and calcium were determined by flame photometer; total nitrogen was determined by Kjeldahl distillation; ammonia was determined by distillation into boric acid; nitrate and nitrite (negligible) were determined by distillation after reduction with DeVarda's alloy. The following ions were determined colorimetrically: Fe^{3+} (thioglycollic acid); Mn^{2+} (potassium periodate); Al^{3+} (ammonium aurine tricarboxylate); and PO_4^{3-} (ammonium molybdate). Ion exchange resins were used to determine the total ionic concentration, the total concentration of strong acids, and Cl^- and SO_4^{2-} . In view of the rapid and uncontrollable changes that can take place in rain water with respect to the total carbonate content, determinations of CO_2 , HCO_3^- , and CO_3^{2-} were omitted.

Dust and the water of Lake Victoria are possible sources of some of the constituents of the Kampala rain water. About 1 kg of dust was collected from floors, tarmac roads, tables, roofs, tree leaves, and other similar places. After

TABLE 1. Major Chemical Constituents, pH, Conductivity, and Lightning Flash Data for Individual Samples of Rain Water Collected at Kampala, Uganda, from August 1, 1958, to July 31, 1959

Date	Precipitation, mm	pH	SO ₄ , ppm	Cl, ppm	NO ₃ , ppm	N, ^a ppm	NH ₃ , ppm	Na, ppm	K, ppm	Total Ions, meq./liter	Strong Acids, meq./liter	Conductivity, ohm ⁻¹ cm ⁻¹	Lightning, no./24 hours
8/12/58	6.5	nil ^c	2.9	...	0.66	0.26	0.11	...	10
8/19/58	10.9	...	1.3	0.4	1.2	...	0.51	0.10	0.06	...	90
8/23/58	17.8	5.9	...	nil	1.1	...	0.68	nil	0.2	0.16	0.07	10	65
8/27/58	4.3	...	9.2	nil	3.6	1.5	0.61	0.6	0.5	0.30	0.20	30	10
9/2/58	17.8	2.2	1.8	1.22	0.4	390
9/3/58	0.5	160
9/4/58	1.3	1.6	1.0	1.2	0.75	0.2	0.2	0.12	0.09	10	15
11/11/58	0.45	10
11/12/58	15.2	5.8	...	0.4	...	1.3	...	0.6	0.2	0.12	0.06	50	30
11/13/58	22.8	7.2	nil	0.6	1.4	0.2	0.39	0.2	0.1	0.20	0.03	15	110
11/15/58	7.9	7.3	nil	0.6	1.8	0.3	0.51	...	nil	0.12	0.03	7	95
11/17/58	1.21
11/18/58	5.8	7.2	nil	0.4	3.0	0.2	0.80	0.6	18.8	0.14	0.04	35	265
11/19/58	nil	...	1.7	0.1	...	0.2	3.4	12	215
11/23/58	4.1	8.1	0.6	0.4	1.7	1.2	0.75	...	1.8	0.14	0.05	12	140
11/25/58	34.8	7.6	0.1	0.5	2.8	1.8	0.77	...	nil	0.10	0.05	7	115
11/27/58	8.6	7.7	1.1	0.4	2.1	1.4	0.46	0.2	0.3	0.14	0.06	10	110
11/28/58	0.7	7.8	2.7	...	0.6	0.8	37	225
11/28/58	1.8	7.7	2.6	...	0.8	3.0	37	230
11/29/58	12.7	7.7	2.5	...	0.8	6.0	37	240
12/3/58	...	7.7	nil	3.1	16.4	3.3	1.55	0.9	9.0	0.32	0.19	32	0
12/8/58	200	245
12/10/58	3.5	5.7	2.8	0.8	1.5	1.0	0.55	nil	2.1	0.16	0.10	12	25
12/11/58	...	8.5	4.8	0.5	1.5	1.4	0.63	0.2	2.8	0.44	0.13	32	90
12/13/58	nil	nil	33	110
12/14/58	...	8.1	1.6	nil	1.0	1.2	0.14	nil	...	0.06	0.02	3	110
12/15/58	5.1	7.0	1.4	nil	1.8	1.4	0.67	0.2	0.4	0.08	0.05	5	400
12/17/58	2.5	6.9	...	4.3	...	3.4	...	9.2	4.1	0.90	0.29	70	50
12/18/58	34.8	8.6	...	0.2	1.5	1.1	0.43	1.6	0.5	0.26	0.02	12	30
12/20/58	15.2	...	nil	0.4	1.4	1.1	0.50	0.4	0.3	0.12	0.03	7	100
12/22/58	3.8	3.0	...	50.8	4.6	170	110
12/25/58	8.4	...	0.3	1.2	2.0	2.0	0.94	203.1	0.6	0.18	0.07	15	100
12/28/58	6.8	...	203.1	12.5	150	140
1/5/59	...	7.6	3.7	...	203.1	16.5	350	0
1/11/59	3.8	8.3	...	7.1	...	2.6	...	40.0	54.0	2.84	0.71	200	16
1/15/59	...	7.9	...	8.1	...	3.2	...	34.0	5.6	2.50	0.68	190	45
1/24/59	8.4	6.6	6.6	6.3	4.0	2.5	0.88	1.6	4.4	0.38	0.25	30	6
1/25/59	7.1	6.8	0.7	0.4	7.4	0.5	0.39	0.1	0.3	0.24	0.04	5	13
2/6/59	10.7	6.6	6.1	0.8	0.2	1.0	0.82	0.6	3.4	0.26	0.15	17	45
2/9/59	1.5	7.6	8.2	1.1	2.9	1.4	0.88	12.8	3.6	0.96	0.25	80	27
2/22/59	8.9	9.0	22.2	6.8	1.1	2.3	0.56	12.3	72.0	3.22	0.66	250	40
2/24/59	1.5	7.6	nil	1.3	1.4	0.8	0.36	0.4	1.7	0.08	0.11	20	8
2/25/59	...	6.8	1.8	6.5	5.9	2.8	2.32	1.2	3.7	0.06	0.05	52	11
2/28/59	14.5	8.3	1.0	0.3	0.8	0.9	0.43	0.1	0.2	0.08	0.04	7	6
3/3/59	1.0	8.6	2.8	0.4	0.3	0.8	0.39	0.4	0.6	0.66	0.07	54	2
3/4/59	...	8.8	1.3	0.9	1.3	...	0.83	7.3	1.2	0.28	0.06	20	2
3/19/59	12.4	8.1	14.5	2.9	7.6	9.6	0.25	159.0	234.0	5.50	0.55	450	16
3/20/59	...	8.9	22.8	0.01	nil	28.0	...	170.0	28.4	11.20	2.48	990	7
3/23/59	0.8	9.8	0.3	...	1.5	1.3	0.48	1.6	0.6	0.20	0.04	15	3
3/24/59	1.8	9.3	36.2	1.4	2.1	2.3	0.97	35.4	3.4	1.90	3.43	170	2

^a N, total nitrogen, by Kjeldahl distillation.^b Within a 40-km radius of Kampala.^c None found.

TABLE 1. Continued

	Pre- cipita- tion, mm	pH	SO ₄ , ppm	Cl, ppm	NO ₃ , ppm	N, ^a ppm	NH ₃ , ppm	Na, ppm	K, ppm	Total Ions, meq./ liter	Strong Acids, meq./ liter	Con- duc- tivity, ohm ⁻¹ cm ⁻¹	Light- ning Flashes, ^b no. in 24 hours
59	0.2	9.3	2.6	2.2	2.1	2.7	0.59	14.6	1.5	1.34	0.24	11	26
59	2.0	8.2	0.9	0.3	1.9	1.4	0.80	0.8	1.6	0.22	0.05	15	25
9	...	9.4	68.9	0.8	6.0	nil	1.34	11.7	18.5	8.80	1.86	745	0
9	...	9.8	46.0	6.5	5.2	nil	nil	140.0	690.0	8.00	1.29	690	10
9	3.8	9.8	20.4	3.8	1.5	1.0	1.03	29.0	157.0	1.50	0.54	157	10
59	0.5	8.2	39.1	3.0	8.5	3.0	1.03	41.4	8.6	3.60	1.11	303	450
59	...	9.1	6.9	2.0	1.7	1.2	0.73	8.8	1.0	0.72	0.22	60	5
59	12.7	8.2	1.6	0.3	1.5	1.2	0.71	1.6	0.3	0.34	0.65	20	30
59	2.0	8.9	1.4	2.8	5.8	1.5	0.93	2.2	1.8	0.76	0.21	67	30
59	0.2	8.0	1.9	0.2	1.3	1.0	0.78	4.2	0.9	0.34	0.07	25	25
59	...	9.2	9.0	2.5	2.2	2.4	2.85	14.0	4.0	1.46	0.30	37	115
59	10.2	9.0	1.8	0.2	0.7	1.0	0.32	1.2	0.3	0.20	0.06	12	40
9	10.7	7.2	nil	0.1	1.5	...	0.44	0.8	0.5	0.26	0.01	17	120
9	1.3	7.9	2.0	1.7	1.4	2.0	0.84	9.6	6.4	0.42	0.14	70	0
9	...	8.2	2.1	2.3	0.8	1.7	0.57	4.2	1.8	0.46	0.13	40	170
59	1.8	7.8	1.7	2.1	1.7	1.4	0.34	2.0	0.9	0.30	0.14	25	15
59	3.0	8.6	1.0	1.2	1.6	1.1	0.53	0.8	30.0	0.36	0.07	30	20
59	2.3	6.6	1.2	0.7	1.0	1.0	0.51	nil	0.4	0.14	0.06	10	5
59	3.8	6.8	2.2	2.5	1.5	1.6	1.00	1.9	1.3	0.32	0.14	27	15
59	...	7.2	0.8	1.8	1.4	1.4	0.56	1.8	0.8	0.28	0.09	20	2
59	...	6.8	7.8	3.3	1.8	3.8	1.58	9.0	3.2	0.82	0.30	72	55
59	...	6.8	nil	0.4	1.8	1.4	0.34	3.6	0.4	0.18	0.04	7	65
59	4.8	6.4	1.8	0.1	1.2	1.5	0.49	nil	0.3	0.12	0.06	7	280
59	...	7.0	1.8	0.2	1.5	1.0	0.46	8.0	0.6	0.14	0.07	12	15
59	...	7.6	4.6	2.2	1.7	0.7	0.27	5.3	1.5	0.52	0.17	45	5
59	20.3	8.6	2.5	1.8	1.1	0.7	0.63	0.8	nil	0.28	0.12	20	25
59	3.6	7.6	4.7	1.9	2.2	1.0	0.83	3.6	6.0	0.50	0.16	45	10
59	22.9	8.0	4.3	1.3	4.0	0.6	...	4.4	12.0	0.54	0.19	40	25

ough mixing, a portion of the dust sample placed in distilled water and shaken in an automatic shaker for 24 hours. The water was filtered and analyzed by the same methods for rain water. A sample of the water from Victoria was also analyzed.

total of 78 rain-water samples was analyzed during the year. The major constituents of these samples, together with their pH values and conductivity, are given in Table 1.

Discussion of results. It is shown in Table 1 that there is a considerable variation in the amounts of the individual major constituents in rain-water samples. The highest amounts recorded are, in ppm: SO₄, 68.9; Cl, 8.1; NO₃, 6.0; total N, 28.0; NH₃, 1.58; Na, 203.1; K, 18.5. The pH values range from 5.7 to 9.8, the highest figure being extremely high for rain

Table 2 gives the median composition of the

Kampala rain water, and the total amount of Na, K, NH₃, total N, NO₃, Cl, and SO₄ deposited per hectare per year. The figures for NH₃ and NO₃ are noticeable as being among the highest so far recorded [Eriksson, 1952a, 1955]. Although high NH₃ and NO₃ figures are generally thought to occur in rain in cities and industrial areas [Gorham, 1958], Junge [1956] expressed the view that these places are not great sources of them. Kampala cannot be considered as an industrial town in regard to atmospheric pollution.

The SO₄ content of the rain is also relatively high. Conway [1942] suggests that in nonindustrial areas an appreciable amount of SO₄ is provided by the oxidation of H₂S to H₂SO₄ in muds. As about 6 per cent of the area of Uganda consists of swamps, this would have been a possible explanation if H₂S were not an uncommon gas in these swamps [Visser, 1960], and if the SO₄

TABLE 2. Median Composition of Kampala Rain Water
(In ppm except where noted otherwise.)

Component	ppm	Amount Precipitated per Year, kg/hectare
pH	7.9	
Na ⁺	1.7	21.7
K ⁺	1.7	21.6
Ca ²⁺	0.05	
Fe ³⁺	0.05	
Al ³⁺	0.04	
Mn ²⁺	0.01	
NH ₄ ⁺	0.63	8.0
Total N	1.5	19.1
NO ₃ ⁻	1.7	21.6
Cl ⁻	0.9	11.4
SO ₄ ⁻	1.8	22.8
PO ₄ ⁻	0.005	
Total ions, meq/l	0.27	
Strong acids, meq/l	0.11	
Conductivity	51	

were not negatively correlated with the pH (see later discussion).

The analytical data were used to examine the correlations with one another, and with the amount of precipitation, wind direction, and number of lightning flashes. To determine the possible correlation between the data the following method was used. Occurrences of quantities in excess of the median value, when found together, were scored +1, and similarly occurrences of quantities below the median, when found together, were scored -1. Occasions when one quantity in excess of the median value occurred together with a deficiency of the other quantity were scored -1. In most cases 78 paired determinations were available, so that, for an uncorrelated pair of measurements, the score was distributed with mean zero and standard deviation $0.5\sqrt{78} = 4.4$. The assigned quotients of scores by the corresponding standard deviations are given in Table 3.

In estimating the significance of the results it should be remembered that on the slightly inaccurate assumption that Table 3 contains 78 independent reduced scores, the level of significance for the highest score, selected because it is the highest, is thereby approximately multiplied by 78. A maximum score of absolute value 4 in Table 3 therefore enables one to conclude

that the table contains a correlation, not due to chance, at a significance level of $78 \times 3 \times 10^{-8} = 2.34 \times 10^{-6}$.

From Table 4 it can be seen that during a shower most ions are washed out and decrease in concentration in the rain water. The same tendency for concentrations of ions to decrease can be observed in the successive rains during a rainy season, although to a lesser extent.

From the spread of the correlation factors (Table 3) it is evident that during a shower the concentrations of the various ions do not all decrease to the same extent. Sodium and chloride ions decrease most rapidly during a shower. The more rapid fallout of NaCl does not support the theory of the differential separation of atmospheric salts by rain [Emanuelson, Eriksson, and Egner, 1954]. The almost constant amount of NO₃⁻ could be thought to support the theory of the formation of nitrates by lightning flashes if it were not for the fact that NO₃⁻ is only very weakly correlated with the number of flashes. If lightning did contribute to the NO₃⁻ content of the rain water, it was due to a minor extent.

Attention has often been drawn to a possible correlation between nitrate and sulfate in rain water; Gorham [1958], for example, found nitrate contents amounting on an average to only under 10 per cent of sulfate. As an explanation of this correlation it was then assumed that lightning occurred as a result of air pollution by combustion. Such a correlation does not exist in the Kampala rain water. On the contrary, sulfate is well correlated with all the cations and anions considered except nitrate and ammonia. It follows from the very reasonable correlations between nitrate and ammonia and between nitrogen and potassium that it is likely that nitrate in the rain water is mainly of organic origin (pollen, microorganisms, ash particles from wood and grass fires). The lack of correlation with sulfate is probably due to the fact that the sulfur originates mainly from another source.

Comparison of the average composition of rain water with that of water that has been in contact with dust is shown in Table 5. The rain water shows a lower sodium content, but a higher potassium, total nitrogen, and sulfate contents than average rain water.

The water of Lake Victoria is another

TABLE 3. Statistical Treatment of Data, quotients of scores by the corresponding standard deviation

	Rain- fall	Flashes	Conduc- tivity	Total Ions	pH	Na ⁺	K ⁺	NH ₃ ⁺	Total N	Strong Acids	NO ₃ ⁻	Cl ⁻	SO ₄ ⁻
Rainfall													
Flashes	+5.32	+5.32	-3.15	-5.90	-2.04	-4.95	-3.47	-1.85	-1.48	-4.48	+0.62	-6.03	-2.88
Conductivity	-3.15	+0.47	+0.47	-6.20	-3.84	-3.13	+1.69	-1.04	+4.71	-3.28	+1.05	-4.00	-4.59
Total ions	-5.90	-6.20	+8.81	+8.81	+1.79	+7.78	+11.2	+4.58	+7.50	+9.12	+3.21	+7.10	+8.51
pH	-2.04	-3.84	+8.81	+8.81	+6.08	+9.32	+7.16	+3.50	+5.35	+12.8	+1.85	+8.94	+9.22
Na ⁺	-4.95	-3.13	+1.79	+6.08	+3.68	+3.68	+2.34	+1.13	-1.07	+4.04	-2.26	+1.61	+4.63
K ⁺	-3.47	+1.69	+7.78	+9.32	+3.68	+4.72	+4.72	+3.02	+5.59	+8.60	+3.51	+8.60	+6.78
NH ₃ ⁺	-1.85	-1.04	+4.58	+7.16	+2.34	+3.02	+5.78	+5.78	+6.60	+8.96	+4.12	+7.02	+6.45
Total N	-1.48	+4.71	+7.50	+3.50	+1.13	+3.02	+5.78	+3.39	+3.39	+4.04	+4.89	+2.04	+3.59
Strong acids	-4.48	-3.28	+9.12	+12.8	+4.04	+8.60	+8.96	+4.04	+6.42	+4.04	+4.20	+6.41	+4.29
NO ₃ ⁻	+0.62	+1.05	+3.21	+1.85	-2.26	+3.51	+4.12	+4.89	+4.20	+3.50	+3.50	+10.0	+10.6
Cl ⁻	-6.03	-4.00	+7.10	+8.94	+1.61	+8.60	+7.02	+2.04	+6.41	+10.0	+4.04	+4.04	+0.57
SO ₄ ⁻	-2.88	-4.59	+8.51	+9.22	+4.63	+6.78	+6.54	+3.39	+4.29	+10.6	+0.57	+5.54	+5.54

TABLE 4. Changes in Ionic Concentrations in Rain Water during the Course of a Shower (Components in ppm except those shown otherwise.)

Components	Period of Sampling from Beginning of Shower			
	1-5 min	5-10 min	10-15 min	15-30 min
pH	6.34	6.66	6.32	5.74
Na ⁺	0.04	0.04	<0.01	<0.01
K ⁺	0.50	0.88	0.81	0.50
NH ₄ ⁺	0.54	0.41	0.24	0.22
NO ₃ ⁻	1.24	0.49	0.68	0.74
Cl ⁻	0.46	0.21	Trace*	0.21
SO ₄ ⁻	2.30	1.56	1.72	1.66
Total N	1.54	0.77	0.70	0.49
Total ions, meq/l	0.27	0.26	0.22	0.23
Strong acids, meq/l	0.06	0.05	0.05	0.05
Conductivity	7.5	5.0	5.0	5.0

* <0.001.

able source of ions in the Kampala rain water. It is known that large areas of water can affect the composition of rain on the adjoining land [Woodcock and Mordy, 1955; Gorham, 1958]. The average composition of Lake Victoria water is given in column 4 of Table 5. The high content of Na⁺ and Cl⁻ of this water in comparison with that of rain water and of the water from the dust sample is striking.

As a result of an attempt to determine, with the help of Table 4, which of the two sources would contribute the various ions to the rain water, it was found that (Table 5, column 5) the greater part of the Na⁺, Fe³⁺, Cl⁻, and PO₄⁻, is likely to originate in the lake, whereas dust seems to contribute largely to the NH₄⁺, total N, NO₃⁻, and Al³⁺. The high amounts of NH₄⁺ and NO₃⁻ in the rain water suggest that these are probably also derived from an additional source.

A sounder basis for these conclusions is made by using Table 3, because a high correlation factor between two rain water constituents makes it very likely that there is a common source for them. From Table 3 it follows that Na⁺ and Cl⁻ possess a very high correlation factor. The high correlation factor between K⁺ and Cl⁻ indicates that a large amount of K⁺ is also derived from the same source, that is, Lake Victoria. It can also be concluded that the

greater part of the SO₄⁻, with its high correlation factors with Na⁺, K⁺, and Cl⁻, likely originates in Lake Victoria. That the SO₄⁻ in rain water has no connection with industrial pollution or with oxidation of hydrogen sulfide in muds is shown by the positive correlation with pH which shows the simultaneous presence of an excess of bases and not the simple occurrence as H₂SO₄ which otherwise would have been expected.

High correlation factors also make a common source for part of the NO₃⁻, NH₄⁺, and total N likely. If, as a result of the dust analysis, it is likely that the major part of the total N originates in this material, then the various correlation factors suggest that the dust also contributes to the K⁺, Na⁺, and Cl⁻ contents of Kampala rain water. Probably the correlation between K⁺ and NH₄⁺ indicates that both likely originate in organic material.

Another method of investigating the source of some of the ions in the rain water was to find if there was any correlation between the minimum ionic concentrations and the wind direction during the period of rainfall. Table 6 gives

TABLE 5. Comparison of the Composition of Median Rain Water with That of Water Mixed with Dust and with Lake Victoria Water (In ppm except for pH and where otherwise noted)

Component	Median Rain Water	Water Mixed with Dust	Lake Victoria Water	Probable Origin of Component
pH	7.9	9.7	7.2	
Na ⁺	1.7	0.94	5.2	Lake
K ⁺	1.7	3.9	3.4	Dust + Lake
Ca ²⁺	0.05	4.0	2.3	Dust + Lake
Fe ³⁺	0.03	0.07	0.29	Lake
Al ³⁺	0.04	0.27	Trace*	Dust
Mn ²⁺	<0.01	<0.01	<0.01	
NH ₄ ⁺	0.63	0.6	0.15	Dust + Lake
Total N	1.5	4.06	1.11	Dust
NO ₃ ⁻	1.7	1.49	0.61	Dust + Lake
Cl ⁻	0.9	0.82	4.3	Lake
SO ₄ ⁻	1.8	4.8	3.1	Dust + Lake
PO ₄ ⁻	0.005	0.005	0.04	Lake
Total ions, meq/l	0.27	1.08	1.07	Dust + Lake
Strong acids, meq/l	0.11	0.15	0.19	Dust + Lake
Conductivity	31	20	100	

* <0.001.

6. Relation between the Maximum Concentration of Components in Rain Water and Prevailing Wind Direction at the Time of Rainfall

Component	Wind Direction	Source of Component
	SE	Lake
	Most directions	Lake + dust
	E	Dust + lake
	NE	Dust
	N and NW	Dust
Acids	Most directions	Lake + dust
	NE and SW	Dust
	Most directions	Lake + dust
	NE-SE	Lake

Results of this investigation. Lake Victoria east of Kampala, and wind coming from direction is likely to have picked up ions the surface of the lake.

Conclusions. The results obtained from comparing the chemical analyses and the correlation of the individual ions with one another with probable sources, lightning flashes, and direction give the following conclusions. Sodium, chlorine, sulfate, and phosphate are derived principally from Lake Victoria, although dust also contributes to their concentration in the rain water. Ammonia, nitrate, nitrogen, calcium, and alumina originate from dust, although the lake and possibly another unknown source contribute to the content. Potassium is likely to originate in Lake Victoria and in the local dust.

Acknowledgments. The rainfall data and wind records were kindly supplied by the Entebbe Meteorological Service, and the information about lightning flashes was supplied by the Uganda Electricity Board. I am indebted to Dr. C. P. Welter, Reader in the Department of Mathematics, Makerere College, for his guidance in the statistical procedures.

REFERENCES

- Conway, E. J., Mean geochemical data in relation to oceanic evolution, *Proc. Royal Irish Acad. B*, **48**, 119-159, 1942.
- Emanuelson, E., E. Eriksson, and H. Egner, Composition of atmospheric precipitation in Sweden, *Tellus*, **6**, 261-267, 1954.
- Eriksson, E., Composition of atmospheric precipitation, 1, Nitrogen compounds, *Tellus*, **4**, 215-232, 1952a.
- Eriksson, E., Composition of atmospheric precipitation, 2, Sulfur, chloride, iodine compounds. Bibliography, *Tellus*, **4**, 280-303, 1952b.
- Eriksson, E., Airborne salts and the composition of river waters, *Tellus*, **7**, 243-250, 1955.
- Gorham, E., The influence and importance of daily weather conditions in the supply of chloride, sulphate and other ions to fresh waters from atmospheric precipitation, *Phil. Trans. Royal Soc. London, B* **241**, 147-178, 1958.
- Junge, C. E., Recent investigations in air chemistry, *Tellus*, **8**, 127-139, 1956.
- Visser, S. A., A study of the decomposition of papyrus in the swamps of Uganda, Ph.D. thesis, University of London (unpublished), 1960.
- Woodcock, A. H., and W. A. Mordy, Salt nuclei, wind, and daily rainfall in Hawaii, *Tellus*, **7**, 291-300, 1955.

(Manuscript received May 1, 1961;
revised August 1, 1961.)

Tritium Geophysics

W. F. LIBBY

*Department of Chemistry and Institute of Geophysics
University of California at Los Angeles*

Abstract. 1. Rainfall from the northern and southern hemispheres has been analyzed for tritium. During the spring, summer, and early fall of 1959, in the northern hemisphere there was a peak in tritium fallout which was proportional to the peak in Sr^{90} fallout, whereas in the southern hemisphere there was no appreciable increase in either the tritium or the Sr^{90} fallout. It seems clear, therefore, that the peak in the northern hemisphere was due to the Russian nuclear bomb tests of October 1958, which injected large amounts of tritium and Sr^{90} into the stratosphere. The proportionality of the two fallout curves suggests that single molecules of water and fine particulate matter have the same residence time in the stratosphere and that the mixing of the stratospheric air with tropospheric air followed by rainfall scavenging must be the mechanism for removal of both from the atmosphere.

2. The general rise in the tritium level due to hydrogen-bomb explosions has been found to be proportional to the total energy release, even though the bombs were fired under a variety of conditions. Therefore, rain and ground-water data can be interpolated and extrapolated so as to supply values for the tritium levels where samples were not collected and measured.

3. The tritium method for determining the age of ground water (since it fell as rain) has been applied to several hydrological problems in southern California and elsewhere. The results confirm hydrological expectations in general but give some new detail and entirely independent checks. When more detailed applications have been made, it appears to promise to be an important tool for water management.

4. Dead rains (abnormally low in tritium) were seen three times in 1960 in California. This, plus the proven low tritium concentration of upwelling water, strongly indicates that extensive areas of upwelling exist and that rains of short atmospheric life occur.

EQUALITY OF STRATOSPHERIC RESIDENCE TIMES FOR FALLOUT PARTICLES AND WATER MOLECULES

Introduction. The method of analysis for tritium in water described earlier by our group [Grosse, Johnston, Wolfgang, and Libby, 1954; Kaufmann and Libby, 1954; Von Buttlar and Libby, 1955; Begemann and Libby, 1957] and by others [Brown and Grummitt, 1956; Berman, Bazan, and Kulp, 1958; Suess, 1958a; Libby, 1960; Carlston, Thatcher, and Rhodes, 1960] has been used to measure samples of rain from Massachusetts, West Ireland, California, and the southern hemisphere.

Results and conclusions. Figure 1 shows for the northern hemisphere (samples from E. A. Martell), the variation in tritium content of the atmosphere throughout 1959 and the spring, summer, and early fall of 1960, together with the average northern hemisphere fallout for Sr^{90} during 1959 [Libby and Palmer, 1960]. Figure 2 presents the results for Valentia, Ireland (samples from J. R. O'Brien), with the available California data added.

(Unfortunately, the critical period in 1959 is missing in the California data.) Figure 3 presents all available data for the southern hemisphere. The detailed data are given in Table 2. It is immediately clear that during the spring and summer of 1959 the peaks attained by both the tritium and Sr^{90} fallout rates in the northern hemisphere were proportional to one another and that there was no corresponding fallout peak in the southern hemisphere. The average proportionality constant at Bedford was about 130 T.U. (1 tritium unit, T.U., is 1 T atom per 10^{18} H atoms) per millicurie of Sr^{90} per square mile per inch of rain, or 13 T atoms per Sr^{90} atom. Considering that the fallout in the spring and summer of 1959 in the northern hemisphere [Libby, 1959; Martell, 1959; Libby and Palmer, 1960; Martell and Drevinsky, 1960; Lockhart, 1960] was due largely to the Russian October 1958 test series, this parallelism of the T and Sr^{90} fallout strongly suggests that both the T and Sr^{90} came from the same source—the island of Novaya Zemlya in October 1958, and that the

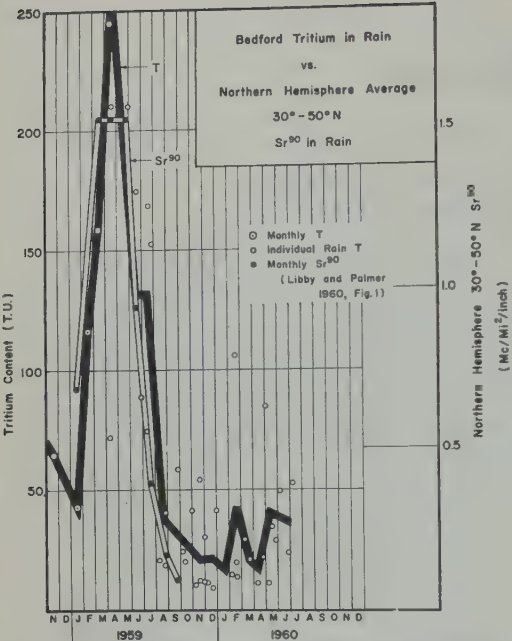


Fig. 1. Bedford tritium in rain vs. northern hemisphere average Sr^{90} in rain.

mechanism of removal from the stratosphere into the troposphere was the same for both T and Sr^{90} . Some of the Bedford rains were measured for Sr^{90} as well as T, giving the results in Table 1 (the Sr^{90} data were furnished by E. A. Martell and P. J. Drevinsky of the Air Force Cambridge Research Laboratory).

The scatter is such that, if the correlation shown in Figure 1 were not so clear, it would be difficult to justify the whole consideration of the T/ Sr^{90} atomic ratio on the Bedford data alone. This indicates that the mechanism by which rainfall precipitates tritium in the troposphere is different from that for Sr^{90} , a very reasonable

TABLE 1. T/ Sr^{90} Ratio at Bedford, Massachusetts, 1959

Period of Collection	Sr^{90} , mc/mi ² /in.	T.U.	T/ Sr^{90} Atoms
Feb. 24-Mar. 31	0.81 ± 0.08	159 ± 4	19
Apr. 1-Apr. 3	0.83 ± 0.08	292 ± 15	33
May 25-June 3	0.17 ± 0.01	220 ± 7	123
July 7-July 13	0.06 ± 0.02	74 ± 7	117
Average 73			

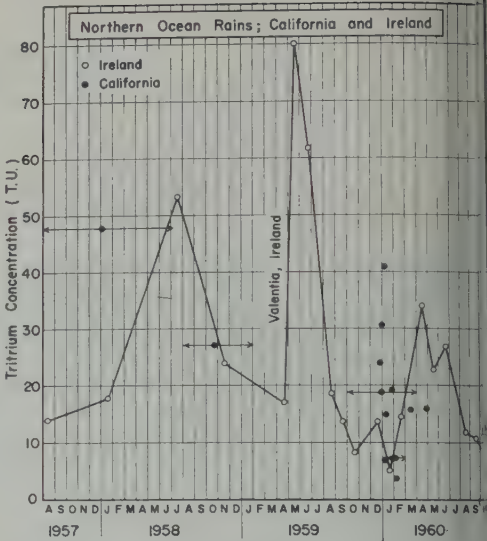


Fig. 2. Northern ocean rains in Ireland California.

result considering the particulate nature of Sr^{90} carrier dust.

The mechanism for the downward mixing stratospheric air must be such as to treat vapor carrying HTO (water in the stratosphere most probably is mainly vapor), and the particulate matter carrying Sr^{90} , in essentially the same way. The finding that molecules and particles have the same residence time in the stratosphere is not unexpected, and all present theories of the fallout mechanism [Stewart, Osmond, Crooks, and Fisher, 1957; Macdonald, 1959; Libby and Palmer, 1960; Feely and Sperry, 1960].

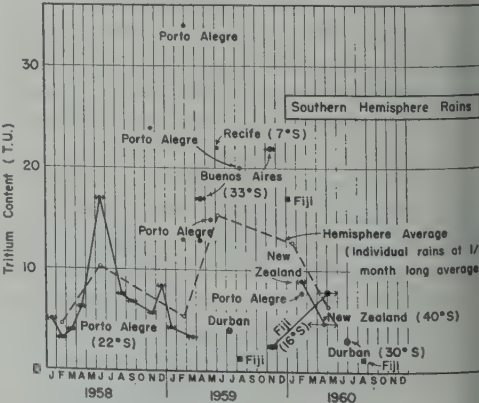


Fig. 3. Southern hemisphere rains.

TABLE 2. Rain Levels

Sample, Source, Collector, and Comments	Date	Tritium Units
United States		
016 Chantilly Road, Bel Air, Los Angeles, Calif. V. F. Libby.	Dec. 20, 1959	24 ± 1.0
Values certainly agree with other United States	Dec. 24, 1959	30 ± 6
East rains for this period (cf. Adam Well, Santa	Jan. 9, 1960	41 ± 5
River recharge stage of winter 1959-60)	April. 27, 1960	16 ± 1
U.C.L.A. campus. By W. F. Libby	Jan. 26-27, 1961	14 ± 0.7
Canyon View Drive, Brentwood. By Anneliese	Jan. 26-27, 1961	19 ± 1.6
er	Mar. 15 1961	16 ± 0.6
Reservoir behind San Gabriel Dam, from canyon	Mar. 1-May 1, 1958	48 ± 2.2
ished. By Los Angeles County Flood Control Dis-	Jan. 5-Apr. 26, 1959	16 ± 1.3
trict, Charles Green of the Foundation and Testing		
ision. (Submitted by same.)		
(rainfall level for the winter season of 1957-58 is		
able considering the sizable H-bomb tests con-		
in August, September, and October 1957 and		
in February 1958. The values for the winter of		
59 are perhaps low compared with those observed		
ford, Mass., and are more in line with those found		
entia, Ireland, also exposed to the open sea.		
County Building, Orange County. By M. R.	Dec. 24, 1959	19 ± 2
erty, Water Resources Center, U.C.L.A. (Sub-		
mitted by same.)		
Monroe Canyon Gauging Station of San Dimas	Dec. 3, 1959	9.1 ± 1.7
erimental Forest, east Los Angeles. (Submitted	Jan. 10-Feb. 28, 1960	7.5 ± 0.6
M. R. Huberty.)		
La Jolla. By W. F. Libby	Nov. 12, 1960	5.0 ± 1.4
Hope Ranch, Santa Barbara. By Dr. V. L.	Jan. 9, 1960	14 ± 0.7
derhoof. (Submitted by same.)	Dec. 2, 1960	6.9 ± 1.5
Five miles from Jolon, Calif., neighboring Hunter	Nov. 26, 1960	2.8 ± 0.5
gett Military Reservation on road to King City.		
W. F. Libby.		
is very low. It appears as though this rain must		
een from an upwelling area in the ocean, and have		
tle time aloft.		
Nacimiento Water. This water was rain collected	Nov. 19, 1959	21 ± 1
ly during the preceding winter with perhaps some	Nov. 27, 1960	10 ± 0.5
over from the rain of the preceding year. By		
nce Taylor. (Submitted by same.)		
i surface water. Island of Molokai. By Terrell L.	Aug. 2, 1960	3.1 ± 1.6
isinger, Land Study Bureau, University of Hawaii.		
mitted by same.)		
Bedford, Mass., By E. A. Martell, Air Force	Nov. 1958	65 ± 11
bridge Research Center. (Submitted by same.)	Jan. 1959	43 ± 6.6
Bedford tritium fallout curve from the weapons	Feb. 1959	116 ± 3
n 1958 closely follows that for northern hemisphere	Mar. 1959	159 ± 4
ing to a peak in April and falling away to the	Apr. 1, 1959	292 ± 15
wide rate by August or September. The Sr ⁹⁰ rates	Apr. 9-13, 1959	72 ± 1.6
uries/square mile/inch of rain) for the 30°-50°N	Apr. 13-20, 1959	220 ± 2.7
or 1959 were:	May 25-June 3, 1959	220 ± 7
0.7	June 15-June 18, 1959	174 ± 2.3
0.7	June 23-26, 1959	89 ± 1.3
1.6	July 7-July 13, 1959	74 ± 7
1.6	July 14-15, 1959	168 ± 4.3

TABLE 2. Continued

Sample, Source, Collector, and Comments	Date	Tritium Units
May 1.5	July 16-20, 1959	152 ± 3
June 0.8	Aug. 6-10, 1959	21 ± 3
July 0.4	Aug. 25-31, 1959	41 ± 5.3
Aug. 0.2	Sept. 29-Oct. 2, 1959	59 ± 1.1
Sept. 0.1	Oct. 2-8, 1959	24 ± 2.1
Oct. 0.1	Oct. 8-9, 1959	20 ± 1.2
The proportionality constant was about about 1 mc/Sr ⁹⁰ /mi ² /in. = 130 T.U.	Oct. 23-26, 1959	9 ± 1.5
	Oct. 27-Nov. 2, 1959	42 ± 4.2
	Nov. 5-9, 1959	11 ± 6
	Nov. 9-16, 1959	12 ± 1.0
	Nov. 16-19, 1959	55 ± 1.4
	Nov. 19-25, 1959	12 ± 1.2
	Nov. 25-30, 1959	31 ± 3.2
	Nov. 30-Dec. 7, 1959	12 ± 1.2
	Dec. 21-28, 1959	9.8 ± 0.8
	Dec. 28-29, 1959	42 ± 1.7
	Dec. 29, 1959-Jan. 4, 1960	16 ± 1.4
	Feb. 4-8, 1960	15 ± 0.8
	Feb. 8-12, 1960	106 ± 2
	Feb. 12-15, 1960	21 ± 0.8
	Feb. 15-19, 1960	14 ± 0.9
	Feb. 24-26, 1960	13 ± 1.8
	Mar. 7, 1960	30 ± 0.7
	Mar. 18-31, 1960	21 ± 1.2
	Apr. 1-6, 1960	11 ± 2.0
	Apr. 20-25, 1960	23 ± 1
	Apr. 28-May 2, 1960	85 ± 5.5
	May 16-24, 1960	29 ± 1.0
	May 24-June 6, 1960	37 ± 1.4
	June 1-6, 1960	50 ± 29
	June 26-28, 1960	24 ± 3.6
Foreign		
Individual rain. Recife, Brazil. Sampled for Health and Safety Laboratory of U.S.A.E.C. By John Harley. (Submitted by same.)	May 2, 1959	22 ± 1
Individual rains. Porto Alegre, Brazil. Combined into monthly totals for Air Force Cambridge Research Center. By Jose Luis Paranhos de Araujo, Sr., I.R.M. Coussirat Araujo, Porto Alegre. (Submitted by E. A. Martell of Air Force Cambridge Research Center, Bedford, Mass.)	Jan. 1958	5.0 ± 0.
	Feb. 1958	3.2 ± 0.
	Feb. 25-Mar. 27, 1958	4.0 ± 0.
	Mar. 29-Apr. 26, 1958	6.3 ± 0.
	June 1958	17 ± 0.
	Aug. 1-Sept. 5, 1958	7.6 ± 0.
	Sept. 6-30, 1958	6.9 ± 0.
	Oct. 25-Dec. 2, 1958	5.8 ± 0.
	Nov. 1958	24 ± 3
	Dec. 1958	8.5 ± 0.
	Jan. 1959	4.4 ± 0.
	Feb. 25, 1959	13 ± 0.
	Feb. 26-Mar. 3, 1959	34 ± 1.
	Mar. 1959	3.5 ± 0.
	Apr. 1959	13.3
	May 1959	15 ± 1.
	Aug. 1959	20 ± 0.
	Feb. 1960	7.8 ± 0.
Rain. Buenos Aires, Argentina. For Health and Safety Laboratory of U.S.A.E.C., New York City, By John Harley. (Submitted by same.)	Apr. 1959	17 ± 1.
	Nov. 1959	22 ± 2

TABLE 2. Continued

Sample, Source, Collector, and Comments	Date	Tritium Units
Considering that this site is on the eastern shore and considerable re-evaporation may have occurred, the values are in line with the sea-water level in the southern oceans at less than 10 T.U. (Samoa sea water; Ohakea, New Zealand, and Lanthala, Fiji, rain.)		
Collected monthly. Valentia Observatory, Glenties, Ireland. (Submitted by J. Byrne, Secretary of National Committee for Geodesy and Geophysics, Dublin.)	Aug. 1957 Jan. 1958 July 1958 Nov. 1958 Apr. 1959 May 1959	14 \pm 0.9 18 \pm 3 53 \pm 1.5 24 \pm 0.8 17 \pm 1.5 80 \pm 20 79 \pm 3
These levels look like California rains, although some- what lower, reflecting the lower tritium level of the Pacific as compared with the Pacific.	June 1959 Aug. 1959 Sept. 1959 Nov. 1959 Dec. 1959 Jan. 1960 Feb. 1960 Apr. 1960 May 1960 June 1960 Aug. 1960 Sept. 1960 Oct. 1960 Nov. 1960 Dec. 1960	61 \pm 2 19 \pm 1.4 14 \pm 3 8.1 \pm 1.6 14 \pm 1.3 5.0 \pm 0.7 15 \pm 1.0 34 \pm 2.1 23 \pm 1.2 27 \pm 1.2 12 \pm 0.7 11 \pm 0.8 5.5 \pm 0.4 4.2 \pm 1.2 16 \pm 1.3
Meteorological Observatory at Ohakea, New Zealand. By E. A. Monk for Health and Safety Laboratory, U.S.A.E.C., New York, John Harley. (Submitted by same.)	Jan. 28-Mar. 31, 1960 Apr. 1-July 1, 1960	9.1 \pm 1.0 4.8 \pm 0.4
These values seem to be low as compared with Fiji on Jan. 1, 1959, at 17.		
Fiji Islands. By Meteorological Office at Lanthala, Fiji. (Submitted by John Harley, Health and Safety Laboratory, U.S.A.E.C., New York City.)	Aug. 6, 1959 Nov. 1959 Jan. 1, 1960 Apr. and May, 1960 Aug. 11, 1960	1.5 \pm 0.4 2.7 \pm 0.7 17 \pm 1.9 7.9 \pm 0.74 1.1 \pm 0.6
Laurenço Marquez, Mozambique. By John Harley Health and Safety Laboratory of U.S.A.E.C. (Submitted by same.)	Sept. 14, 1959	1.5 \pm 0.6
This sample seems to show very little except cosmic-ray activity.		
Durban, South Africa. By John Harley for Health and Safety Laboratory of U.S.A.E.C. (Submitted by same.)	July 3-12, 1959 June 23-24, 1960	4.1 \pm 0.3 3.2 \pm 0.8

implicitly accept it. The physical basis is
no appreciable settling of the fine particu-
les involved in stratospheric fallout occurs as
a result of gravity. For all intents and pur-
poses the particles move with the air mass just
as the true gases do; therefore the expectation
is that all radioactive products of nuclear

explosions should show the same fractional rate
of downward transport from the stratosphere,
and that, except for the complication of re-entry
into the stratosphere for the long-lived gaseous
products that have no appreciable surface
hold-up, there would be essentially no distinc-
tion between any of them.

TABLE 3. United States River and Lake Water; Ocean Surface Water

Sample, Source, Collector, and Comments	Date	Tritium Units
Water. Colorado River. By Dr. H. E. Pearson of Metropolitan Water District.	Feb. 1960	52 ± 1.4
Water. Mississippi River, at St. Louis. By Dr. A. C. Wahl. (Submitted by same.) This value, though high, is not so high as expected. Further measurements are indicated.	Oct. 16, 1960	119 ± 2
Water. Lake Michigan. By A. F. Hakanson, Chemistry Department, University of Chicago. This is to be compared with the prebomb cosmic ray tritium level of 1.64 ± 0.04 [Kaufman and Libby, 1954] and the post-Castle 7.5 T.U. Taking the re-evaporation downwind effect, which probably amounts to about a factor of 3 as judged from the cosmic-ray data, the Castle ocean rise should have been 5.6/3 to about 3 T.U. from 1, and the present northern ocean level should be about 9 T.U., which appears to be an acceptable value.	Feb. 1, 1960	28 ± 3
New (Dry Niagara) wine. Grapes grown at Naples, western New York State. By Widmer's Wine Cellars. (Submitted by William Widmer.) This looks about right for ground water at this locality in the spring and summer of 1960 and is to be compared with the 1940 to 1952 value of 6 T.U.	1960 vintage	90 ± 2 97 ± 1
Beach water. Pacific Ocean at Santa Monica, Calif. By Gerry Marvray. These are to be compared directly with the sample collected at the same spot on June 8, 1953, which gave 0.54 ± 0.02 T.U. [Von Buttlar and Libby, 1955]. The low value in 1961 is presumably due to upwelling (see section 4).	Mar. 9, 1960 Apr. 4, 1961	14 ± 2.3 2.7 ± 1
Surface water. Atlantic City Pier, Atlantic Ocean (depth 24 ft). By Peter Simpriano. (Submitted by O. R. Angelillo.) This value may reflect the fact that rain from the eastern shore is particularly high in tritium.	End Apr. 1960	23 ± 1.3
Surface sea water. Johnson Island, in Central Pacific, 3 miles offshore (depth approx. 30 ft). (Submitted by H. E. Suess, La Jolla.) This value is so high that it shows incomplete horizontal mixing in a matter of 2 years—the time since the last surface H bombs in the Pacific. These surface shots produce surface water of about 20,000 T.U. over an area of perhaps 1000 square miles (unpublished data of the author's on the Navajo test in 1958). This level could correspond to a present area of about 1,000,000 square miles, or a good fraction of the Central Pacific. It would be of the greatest interest to trace these tritium islands as they grow and to check the results with the Sr^{90} and Cs^{137} , which should be high also.	Feb. 4, 1960	21 ± 2
Surface sea water. Lanthala Bay, Fiji Islands. By New Zealand Meteorological service. This is the highest value obtained so far for southern surface Pacific Ocean water. It may reflect southward movement of local tritium fallout from the bomb tests at Bikini or Eniwetok.	Aug. 11, 1960	$8.6 \pm 1.$

TABLE 3. Continued

Sample, Source, Collector, and Comments	Date	Tritium Units
sea water. Lago Lago, American Samoa, approxi- ly 1½ miles south of Breakers Point 1½ miles from e (depth 300 to 400 ft; light rain falling). (Submitted L. E. Suess, La Jolla.)	Feb. 1, 1960	5.7 ± 1.0 6.4 ± 2.0
South Pacific waters (15°S) seem to be free of the fallout tritium from the surface H-bomb tests at Tok and Bikini (11°N), in contrast to the North (cf. Johnson Island surface water on Feb. 4, 1960, had been 21). This appears to be reasonable, the currents being generally toward the northwest from that area, and considering the general system of clockwise rotation in the northern oceans.		
Sea water. Dead Sea. (Submitted by K. O. Emery, Geology Department, University of Southern California.)	Feb. 1, 1960	8.5 ± 0.3
Sea water. Mediterranean Sea: near Sicily and South Sardinia (39°09'N, 9°42'E); and center of Tyrrhenian Sea (39°59'N, 12°36'E). By Prof. Giuseppina Aliverti, Istituto di Meteorologia e Oceanografia, Istituto Universitario Navale, Naples. (Submitted by same.)	Oct. 19, 1959	5.5 ± 0.6 10.6 ± 1.8

2. THERMONUCLEAR TRITIUM

Production. The tritium content of rain and surface waters has been measured more or less continuously since 1950 (see references cited in introduction, section 1). These data, together with the currently reported rainfall data in Table 2 and further measurements on surface waters given in this section, permit a comparison to be made with the published total thermonuclear energy release [Libby, 1959].

Data and conclusions. Data on the tritium content of rivers and lakes in the United States and of the oceans are given in Table 3. In Figure 4, these data are plotted, together with the data cited above. From Figure 4, we see that the rise in tritium content in post early fallout surface waters or land rains (3 to 6 months after the tests) is roughly proportional to the estimated total number of megatons of very explosive energy equivalent of thermonuclear fusion energy released [Libby, 1959], despite the wide range of firing conditions with frequent wide variations in local quick fallout tritium into the ocean. This is particularly important in respect to some of the tests in the Pacific, where the fireball touched the ocean surface.

Interestingly, in the bomb tritium fallout data of the features of the cosmic-ray data are

preserved. Among these are the ratio of tritium in inland (Midwest) rain to that in California rain. The California rain figure before the bombing was about 2.5 T.U.; after 87 megatons of fusion it was about 20 T.U.—about a tenfold increase. Similarly, Chicago rain rose from 6 to

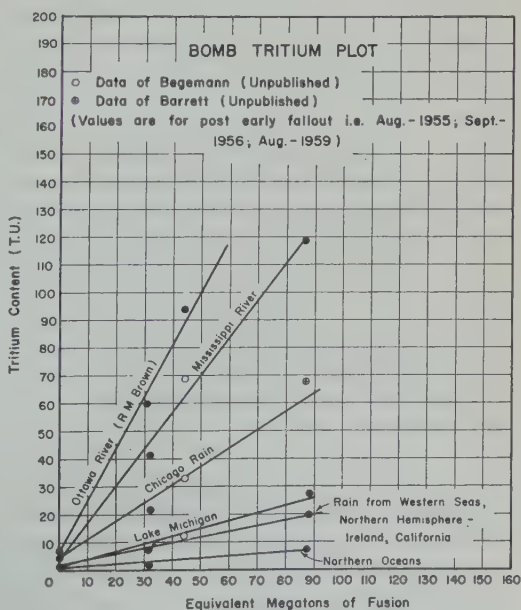


Fig. 4. Bomb tritium plot.

68 T.U. Some of the features of the artificial tritium fallout are different from the cosmic-ray tritium because of the suddenness with which the bomb tritium was injected. For example, Lake Michigan rose from 1.6 to 28, because the great depth of the lake (95 meters) with an effective drainage area about equal to the lake was in full equilibrium with the steady cosmic-ray production, a good part of the tritium in the lake decaying to form He^3 before leaving. And so the mean cosmic-ray production over 18 years is to be compared with a total bomb fallout.

It is of interest that the tritium content of wines grown in western New York State rose from 6 to 90 T.U. as a result of the bomb fallout, in rough general agreement with the rain and ground-water values. Also, the Ottawa River values again are higher than the Mississippi River values, just as was true for cosmic rays. From all these observations, we conclude that there is a very considerable parallelism between the fallout of cosmic-ray-produced tritium in the high stratosphere and that of hydrogen-bomb-produced tritium, which also largely comes from the stratosphere.

3. HYDROLOGY

Introduction

The tritium method for determining the age of underground water [Grosse, Johnston, Wolfgang, and Libby, 1951; Kaufman and Libby, 1954; Von Buttlar and Libby, 1955; Kaufman and Orlob, 1956; Von Buttlar and Wendt, 1958; Brown and Grummitt, 1956; Gilletti, Bazan, and Kulp, 1958; Suess, 1958b; Brown, 1960; Carlston, Thatcher, and Rhodehamel, 1960] has been applied to several hydrological problems, mainly in the southern California area. The data are given in Table 4, and the particular problems are described and discussed below.

The tritium hydrological method is the use of the tritium content of ground water to determine how recently it fell as rain or the ratio of the mixture of rain with old water (older than about 50 years, an age such that the tritium levels will have been reduced to about 5 per cent of the original rain level as the result of radioactive decay, according to the 12-year half-life). It also uses injected tritium to study detailed local hydrological problems, a method,

however, that has not been as widely used as the rain method.

The course of the curve of the rain tritium content with time (see sections 1 and 2) allows an estimate to be made of the probable tritium concentration in rain and snow for most regions, although some further detail may sometimes be needed. The curve has shown a marked change since 1954 (see section 2), owing to explosion of thermonuclear weapon devices in the atmosphere. Thus, recent rains are distinctly labeled as compared with pre-1954, when only cosmic-ray tritium was generally present, and therefore it is possible to identify various natural waters definitely. For example, four general water types are found in southern California: (1) old water which contains no measurable tritium (cf. 1. Sparklett and San Fernando City wells discussed later); (2) fresh rain containing about 20 T.U. (section 1); (3) the Owens Valley water from the Sierra Nevada range, which has about 10 T.U. (Table 4, section on San Fernando) and (4) the Colorado River water, which has about 52 T.U. (section 2).

Consequently, it is possible in principle to follow the water brought into southern California to follow the water brought in by the aqueducts from both the Sierras and from the Colorado River as well as to check the contribution of local rain and the rate of depletion of the very old underground water supplies.

Hydrological Problems

1. *Sparkletts wells* (samples from O. Angellilo and D. J. Finlayson). An area of Eagle Rock near Santa Monica on the southern California coast has artesian wells (the water is sold for drinking purposes by the Sparklett Company). The wells are located either along the fault (Raymond Fault) or upstream in a general sloping aquifer which is exposed to the surface some 2 miles away. The depth of the aquifer along the fault is about 100 feet, and the general thickness of the aquifer lying between its closing clay layers is about 20 feet. The wells along the fault (3987A, 3987B, and 3987F) are in a line 500 feet long and about 250 feet apart. The first two showed essentially no tritium (0 ± 0.26 , 0 ± 1.5 , -0.6 ± 0.6 T.U. for 3987A and -0.17 ± 0.5 T.U. for 3987B); the third in line showed some (4.7 ± 0.5 and 3.5 ± 0.5 T.U. for 3987F). One well 3000 feet upslope

TABLE 4. Hydrological Tritium Data

Description of Sample, Source, and Collection	Date	Tritium Units
Catalina Island		
from Avalon Public Service Well 1 (Riess), Avalon, Catalina Island, Calif. (See description in text.) Collected and submitted by O. R. Angelillo, Los Angeles, and Avalon Public Service Co., W. L. Scott, president. The well water is old, certainly older than 1954, when the first rise in rain tritium levels due to H-bomb testing was observed.	Mar. 30, 1960	2.5 ± 0.8
Sparkletts Wells (Sparkletts Drinking Water Corp., 4500 York Blvd., Los Angeles)		
from Sparkletts Well 1 (Los Angeles Water District 3987A). Well is about 100 ft deep, perforated all the way through a gravel envelope; water comes from a layer covered by a hard layer of blue-black clay. First sample submitted by O. R. Angelillo, Los Angeles. Second sample collected under full artesian flow by D. J. Finlayson, State Water Rights Board; submitted by same. This water appears to be at least 40 years old, calculated from the cosmic-ray tritium level of 2 to 3 T.U. in California which applied prior to 1954 when H-bomb tritium arrived first in appreciable quantities.	Apr. 11, 1960	0 ± 0.26
		0 ± 1.5
		0.3 ± 1.0
	Oct. 10, 1960	-0.61 ± 1.6
from Sparkletts Well 3, Eagle Rock Area (Los Angeles Water District 3987F). Well is about 500 ft east of No. 1, has a total depth of 278 ft, and is perforated from 124 to 129 ft and again from 145 to 151 and from 200 to 266 ft. It lies just upstream from the Raymond fault. This well, Sparkletts Well, and Eagle Rock Area Well 3986B all lie in a simple artesian pressure area and forebay, the lower end of which is defined by the Raymond fault. From the fault upstream approximately 7500 ft to the north, the formation consists of an aquifer overlain by a relatively impervious aquiclude of sandy clays. In the area over 7500 ft northerly, the aquifer is exposed to form a forebay area. First sample collected by Brown and Angelillo and submitted by D. J. Finlayson, of State Water Rights Board. Second sample taken under full artesian flow and submitted by D. J. Finlayson, of State Water Rights Board. This is old water but apparently not as old as that from Sparkletts Well 1, 500 ft to west and also lying on the fault. Rain has been well above 4.7 T.U. in this area since 1954; it appears to be about 6 years old, or, of course, may be a small amount of recent rain in what otherwise is old water.	June 24, 1960	4.7 ± 0.5
	Oct. 9, 1960	3.5 ± 0.6
from Sparkletts Well 2, Eagle Rock Area (Los Angeles Water District 3987B). Well 2 is about equivalent between Wells 1 and 3. It is gravel packed from about 150 ft below ground level. Sampled while under full artesian flow and submitted by D. J. Finlayson, the State Water Rights Board. This water appears to be at least 40 years old, calculated from the cosmic-ray tritium level of 2 to 3 T.U. in California, which applied before 1954, when H-bomb tritium arrived in appreciable quantities.	Oct. 9, 1960	-0.17 ± 0.5

TABLE 4. Continued

Description of Sample, Source, and Collection	Date	Tritium Units
<p>Water from Eagle Rock Well 3986B, on Cauchon property at Ridgeview. Well is 3000 ft northerly of Raymond Fault in pressure area between Fault and surfacing of the aquifer about 4500 ft farther north.</p> <p>This also is old water although, again, apparently not as old as the water from Sparkletts Well 1. Like Well 3, it appears to be about 6 years old.</p> <p>All the water in the aquifer leading to the Raymond Fault is several years old. Some is older as though different rates or recharge in different parts of the aquifer were involved.</p>	June 24, 1960	2.1 ± 0.7
San Fernando City		
<p>Water from City Well 1 at intersection of 4th St. and Hubbard St. Well is 12 in. in diameter and 566 ft deep. It produces 900 gal water/min and is pumped intermittently into the city reservoir. (See description in text.) Sample taken by S. E. Bergman; submitted by A. L. Sonderegger.</p> <p>No recent rain in this water, in particular, no appreciable flow from Lower City of San Fernando Reservoir 1.5 miles to west.</p>	Sept. 30, 1960	$-0.25 \pm$
<p>Water from Lower City of San Fernando Reservoir. Collected by S. E. Bergman; submitted by A. L. Sonderegger.</p> <p>This water is higher than recent rains and would appear to be coming mainly from Owens Valley through the aqueduct.</p> <p>Water from Mission Well 5, about halfway between Reservoir and City Well 1.</p> <p>This water is not reservoir water.</p>	Sept. 30, 1960	36 ± 5
Kern River		
<p>Sample taken from Kern River near Bakersfield, Calif., on Rio Bravo Ranch about $\frac{3}{4}$ mile west of junction with Cottonwood Creek. Kern River is thought to be ultimate source of most of ground water in area sampled by the Standard Oil Well, Amerada Well, and Haddad Well, although some water flows in from Cottonwood Creek. Sample collected and submitted by John C. Manning, Bakersfield.</p> <p>This result indicates that the Kern River at this point has some old water, since the expected value for new rain or snow would exceed 20 T.U.</p>	Apr. 6, 1960	16 ± 0.6
<p>Water from Standard Oil Well, at point 2 miles due south of sampling point of Kern River sample. Well is thought to be producing from both upper Kern River formation water sands and lower Santa Margarita water sands. Some oil present from oil streaks in upper formation. There is thought to be no recharge from surface rain. Sample collected and submitted by John C. Manning, Bakersfield.</p> <p>This result taken with the Kern River value of 16 ± 0.6 T.U. is compatible with the proposed hydrological flow pattern, assuming water from Santa Margarita sands to be too old to contain tritium.</p>	Apr. 6, 1960	$5.7 \pm 0.$

TABLE 4. Continued

Description of Sample, Source, and Collection	Date	Tritium Units
<p>from Haddad Brothers Well about 5 miles south of River and 2 miles southeast of Standard Oil Well. thought to be downstream for underground flow from Wood Creek about 2½ miles northeast. Collected submitted by John C. Manning, Bakersfield. young water, and the well is subject to rapid re- with rain not over 1 or 2 years old.</p>	Apr. 8, 1960	21 ± 1.4
<p>collected from Amerada Well about 2 miles from River on Rio Bravo Ranch. Thought to be down- for ground-water flow from Kern River. Collected submitted by John C. Manning, Bakersfield. well is subject to fast recharge. The water is very perhaps 6 months or 1 year old.</p>	Apr. 8, 1960	31 ± 5
Salinas Valley		
<p>from Salinas River. First sample collected at King by Bruce Taylor; submitted by same. Second sample collected at Bradley.</p>	Nov. 29, 1959	15 ± 0.6
<p>values are lower than for pure rain as shown in Nacimiento (21) and Jolon rain (27). Apparently some and water is involved in the river flow.</p>	Nov. 27, 1960	12 ± 0.4
<p>water from Salinas Valley. First sample from King Municipal Well (Canal St. 2), about ½ mile from River, into which Nacimiento Lake water is dis- ed. Second sample from J. C. Hickey Well, about from Salinas River, and in Pine Canyon, which into Salinas Valley. Third sample from Jolon Well, Jolon Grade from Salinas Valley, about 10 miles Salinas River and 4 miles from San Antonio River, flows into Salinas near Bradley. Samples collected Bruce Taylor and W. F. Libby.</p>	Nov. 29, 1959	2.0 ± 0.3
<p>these wells seem to be old water, at least 6 years old City, and possibly 40 years old in the Hickey Well Canyon and the Jolon Well.</p>	Nov. 29, 1959	0.6 ± 0.3
	Feb. 21, 1960	0.3 ± 1.8
San Bernardino		
<p>water from San Bernardino County Well. Well 829 ft perforated in bottom 500 ft, 3 miles east of Yucaipa, ion Sec33EZ, T1SR1W, 1090 ft east of west line of on 33, and 1410 ft north of south line of Section 33). le taken by Mr. Chandler of San Bernardino micipal Water District; submitted by D. B. Willits Robert Thomas of Southern California District of Department of Water Resources.</p>	Mar. 25, 1960	7 ± 2
<p>water is not recent rain.</p>		
Orange County Recharge		
<p>ge County, in common with many areas in Southern ia, is experiencing an overdraft on its ground water To augment the natural recharge of the Santa Ana water from the Metropolitan Water District (to River) is used for recharge. Four samples were d during March 1960 from Lower Santa Ana Basin recharge effect. Submitted by M. R. Huberty, Resources Center, U.C.L.A.</p>		

TABLE 4. Continued

Description of Sample, Source, and Collection	Date	Tritium Units
Surface water sample from River Trout Farm (74-18-5B), from upper part of alluvial fan at SE $\frac{1}{4}$ of SW $\frac{1}{4}$ of Sec7 T4SR9N. This is not recent rain water.	Mar. 21, 1960	2.2 ± 0.9
Water from Orange Co. District Well 3 (91-37-11A), on mid portion of basin in SW $\frac{1}{4}$ of the SW $\frac{1}{4}$ of Sec34 T48SR10W. This water is younger than 6 years. The general level of rain in this area before 1954 was about 2 T.U.	Mar. 21, 1960	9 ± 2.9
Water from Orange Co. Well (112-10-11A), the J. E. Lecravian Irrigation Well about 2 $\frac{1}{2}$ miles from coast near lower end of basin in NE $\frac{1}{4}$ of SE $\frac{1}{4}$ of Sec36 T5SR11W. This is old water like the River Trout Farm surface water.	Mar. 22, 1960	2.8 ± 0.9
Water from Orange Co. Well (77-13-16A), Yerba Linda County Well near head of alluvial fan about 1 mile from stream channel of Santa Ana River. This water is younger than 6 years. General comment. It appears that recharge is working and is creating islands of new water in an old water table.	Mar. 22, 1960	8.4 ± 1.1
Santa Maria Valley		
Water from Well 10/33-21R1 on L. H. Adam property on floodplain of Santa Maria River. No surface flow at time. Well receives recharge directly from infiltration of river flow during each season of runoff. Collected and sub- mitted by John C. Manning, Bakersfield. Looks right for rain and river water in this area.	Apr. 13, 1960	19 ± 1.2
Water from Gilliland Well (Old Norsking Well 9/33-8K1) on Orcutt upland about 5 miles south and upland from Santa Maria River (T9N-R33W). Bradley Canyon lies between this well and river. It is thought to receive re- charge only from surface infiltration and small ground- water flow. Collected and submitted by John C. Manning, Bakersfield. Both this and Adam Well are described thoroughly in <i>U.S.G.S. Water Supply Paper 1000</i> . Well is fed mainly by ground recharge of water at least 5 or 6 years old on the average.	Apr. 13, 1960	3.2 ± 0.8
Montebello Forebay		
Samples of ground water were taken to provide data for Montebello Forebay of Central Basin of Los Angeles County Flood Control District. Collected and submitted by Charles Green, of L. A. Flood Control District.		
Sample 1600Z from well perforated between 525 and 535 ft in lowermost aquifer of Rio Hondo Spreading Grounds. This aquifer is thought to be below the Silverado. Water level in well 90.81 ft above sea level.		1.6 ± 0.7
Sample 1599D from near Rosecrans Blvd. and the San Gabriel River perforated between 626 and 628 ft. Water level 94.9 ft above sea level. Near top of Silverado Aquifer.		1.9 ± 0.7
Sample 1519J from well near Rosecrans and Atlantic Blvds., perforated between 105 and 127 ft in Gaspar aquifer. Water level 106.8 ft above sea level. Gaspar Aquifer is above the Silverado and just below Los Angeles River.		1.8 ± 0.6 1.8 ± 0.6

TABLE 4. Continued

Description of Sample, Source, and Collection	Date	Tritium Units
<p>sea from which the samples were taken was recharged Colorado River water in the late summer of each year beginning in 1954. The results show the water to be older than the completely mixed recharge in all three samples and to contain Colorado water or recent rain in any appreciable percentage.</p>		
Molokai Dike Complex		
<p>An attempt to establish the approximate length of time water has been stored in an old volcanic eruption with wall components called a dike complex, samples of water in a tunnel on the Island of Molokai have been taken and studied. The 5-mile-long tunnel extends from the dry side of the island to the wet side and was in various stages of construction. It will intercept a stream of water on the wet side of the island and carry it through to the dry side and thus furnish ample water during the winter for irrigation. In summer the stream flow is often low, and the terrain is so rugged that damming the stream is not practical. However, the geological rock formation of the stream is expected to serve as a reservoir; if the dike complex was reached by the tunnel, stored water would be released. The point of the tritium measurement would show whether water from the dike complex is old water or recent rain which traveled by rapid percolation through the dike. If the water proves to be low in tritium, the area would be used as a reservoir to store water during winter for dry summers. If the dike complex does not hold water for a period of several months, 5 or 6 pumps would have to be installed farther down stream and run more or less continuously during the summer to lift water approximately 100 ft to be discharged into the tunnel.</p>		
<p>Sample taken from horizontal hole core drilled 15 ft from tunnel center line at Station 236 + 08 L. Second sample, from springs issuing at tunnel spring line at Station 247 + 50 L; third sample, from Station 244 + 86. Samples collected by Robert T. Chuck, of State of Hawaii Water Authority, and submitted by Terrell L. Wenger, of Land Study Bureau, University of Hawaii. It appears that the second sample at the tunnel spring line have been 5 or 6 years old. It is completely clear, however, that both the first and third samples were older water, older than 6 years, in view of the bomb test tritium since 1954. The present Pacific surface water is 10 to</p>	<p>July 17, 1960 July 17, 1960 June 18, 1960</p>	<p>0.21 ± 0.3 2.8 ± 0.4 0.38 ± 0.3</p>
<p>trifer (3986B) showed 2.1 ± 0.7 T.U. Apparently all this water is at least several years old or is a mixture of a relatively small amount of new rain with water old enough to contain tritium (50 years or so), and some of the waters along the fault and in the aquifer may be different. The water from the first well right on the fault (A and B) appears to be older, or to have less recent rain mixed in</p>	<p>it, while that in the third well (F) and in the aquifer look similar with a calculated mean age of only 6 years or so.</p> <p>2. <i>San Fernando City Well</i> (samples from A. L. Sonderegger). In the City of San Fernando, California, since 1885 a well has been in intermittent daily use. It is 12 inches in diameter and 566 feet deep. The general location was in use as a source of spring water for many</p>	

years before 1885 by Mission settlers and Indians. Old records and maps show springs and artesian wells at the same place. It is 1.5 miles east of the large San Fernando Reservoir, which is filled with Owens Valley rain water (36 ± 5 T.U.). This well water is essentially tritium-free (-0.25 ± 0.3 T.U.). It thus appears that the well water is dissociated from the surface water system and, in particular, that the reservoir water is not contributing to the well-water flow.

The Mission Well about halfway closer to the reservoir from the City Well shows a little rain (3.4 ± 0.5 T.U.) but is still not part of the reservoir water system.

3. *Orange County Recharge* (samples from M. R. Huberty). As the water table near the City of Santa Ana has been depleted by use over the years, a recharge operation has been conducted with Colorado River water (estimated 50 T.U.), and a study of the recharge area has shown a spotty distribution of the recharge water. It seems as though the Colorado water goes into certain depleted areas and compresses and raises the level of the remaining old ground water in others.

4. *Catalina Island* (sample from O. R. Angelilo). The single sample analyzed came from one of the wells (Riess at Avalon) on Catalina Island and was taken with the well casing reaching to 421 feet below sea level. The total depth of the well was 987 feet with a clear depth of 922 feet, and the bottom of the casing was 856 feet below the surface. The result, 2.5 ± 0.8 T.U., shows that the water is old—at least of prebomb age—or is a dilute solution of recent rain with tritium-free water from some old source.

5. *Montebello Forebay* (samples from Charles Green). The 7-year-old recharge operation with Colorado River water in the Montebello Forebay of the Central Basin of the Los Angeles County Flood Control District seems not to have reached the wells sampled to any appreciable extent, and all three aquifers, the one below the Silverado at Rio Hondo (sample 1600Z), the upper Silverado (sample 1599D), and the Gaspar (sample 1519J), consist of water 10 years old or older.

Molokai Dike Complex (samples from Terrill Noffsinger). The vertically stratified rocks in the old volcanic crater on the Island of Molokai, Hawaii, are shown to have water several years

old stored in them. This is of considerable importance to the Hawaii Water Authority, assuring a steady supply on the downwind (leeward) side of the island mountain by a tunnel cutting through to a stream on the wet upwind (windward) side. The permeable compartments banded by volcanic debris can be used for water storage.

Kern River (samples from John Manning). The general pattern of underground flow near the Kern River near Bakersfield, California, appears to be about as expected, with only two exceptions.

Santa Maria Valley (samples from John Manning). The Gilliland Well 5 miles south and upland from the Santa Maria River seems not to draw on the river. A canyon (Bradley Canyon) lies between this well and the river, and so its water was expected to be different from the river, as the data confirm.

Salinas Valley (samples from Bruce Taylor and W. F. Libby). The King City Municipal Well about half a mile from the Salinas River is apparently not in rapid equilibrium with the river, the water in the well being at least 10 years old.

The Hickey Well is farther away and apparently has very old water, as does the Jolon Well which is even farther away. All this strongly indicates that the rate at which the river water recharges the ground-water reservoir is very low, being less than 0.5 mile in 6 years, since water younger than 1954 seems to exist in the King City Well.

Conclusions

The tritium hydrology method continues to look promising as a practical field tool for water management. The present results in general are confirmatory of the expectations of hydrologists but furnish an independent check. They indicate that, with more detailed hydrological study and sampling, important new results can be obtained.

4. LOW-TRITIUM RAINS AND LARGE-SCALE UPWELLING IN THE OCEAN ALONG THE CALIFORNIA COAST

Introduction. The tritium method (see sections 1, 2, and 3) in the course of application to rains and surface ocean water along the California coast has shown regions of low tritium concentration that seem to be explicable on the basis of large-scale upwelling w

TABLE 5. Dead Rains (less than 7 T.U.) and Surface Ocean Waters

Sample, Source, Collector, and Comments	Date	Tritium Units
Jolla, Calif.* By W. F. Libby.	Nov. 12, 1960	5.0 ± 1.4
Pe Ranch, Santa Barbara, Calif. By V. L.	Dec. 2, 1960	6.9 ± 1.5
hoof.		
re miles north of Jolon near King City, Calif. By	Nov. 26, 1960	2.8 ± 0.5
Libby.		
urface water. First samples (7120) just south of	About Aug. 20, 1960	1.8 ± 0.3
Cruz Island ($32^{\circ}50.2'N$; $119^{\circ}42.2'W$);		0.6 ± 0.4
samples (7122) from open sea about 80 miles		1.6 ± 0.2
of first sample ($31^{\circ}52'N$; $120^{\circ}26'W$). By K. O.		0.9 ± 0.2
, Department of Geology, University of Southern		
nia.		
ears that extensive upwelling is occurring in this		
ause the tritium levels are far below other surface		
s in the Pacific (10 to 15 T.U.; see Section 2) and		
e in the world (5 to 10 T.U.; see Section 2). It also		
hat even the sample farthest south is showing		
g.		
Mouth of Salinas River as it empties into ocean.	Nov. 29, 1959	1.1 ± 0.5
ncentration 1.8% vs. 3.4% for sea water. By Bruce		
ater. Breakers at Santa Monica at same spot as in	Apr. 4, 1961	1.5 ± 0.7
1960 (14 ± 2.3 T.U.); June 8, 1953 (0.54 ± 0.02		3.9 ± 1.4
usly, the beach water is much lower than it was in		
dicating upwelling in April 1961. It also appears		
n the sample farthest south is showing upwelling.		

eral average for California rains was about 20 T.U. (see section 1).

areas extended enough to give rains over
able areas in California. The fact that
as arrive low in tritium means also that
a followed by the moisture in these places
cularly short from the ocean to the land
e.

The data are listed in Table 5.
ssion. The sketchy data now in hand
in any way completely delineate the
of the areas of upwelling, but they do
hat beach water (Salinas and Santa
) can have risen recently from the ocean
and that the upwelling situation can
om time to time (Santa Monica). Ap-
y both these results are reasonable to
raphers.

ps more surprising, however, is that the
upwelling are large enough to provide
sture for entire rainstorms, and that the
eric moisture so produced does not stay
ir more than a day or so at most. The
oint rests on an earlier estimate of about
[Von Buttlar and Libby, 1955] from

cosmic-ray tritium studies and the ratio of the
atmospheric pickup for these rains (less than 1
to 2 T.U.) to the normal pickup (6 to 10 T.U.).

The evaporation from upwelling areas raises
an additional difficulty, since presumably the
upwelling waters are considerably colder and
therefore have a lower vapor pressure than the
main surface waters. The short atmospheric
residence time is no great surprise to meteorolo-
gists, although it does give them a new tool for
determining the pathway of water from the
ocean to the land in these cases. It would, of
course, be interesting to see whether these rains
were low in Sr^{90} and Cs^{137} radioactive fallout
from the stratosphere (see section 1), as they
certainly would be expected to be.

Further measurements on dead rains will
probably be a considerable help in settling these
questions. There appears to be little doubt,
however, that there are extensive areas of up-
welling off the California coast and that they are
occasionally the source of considerable rainfall
within 200 or 300 miles downwind.

Acknowledgments. This study was supported by the Water Resources Center of the University of California at Los Angeles, the University of California Research Committee, and the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command. It is particularly pleasant to thank these sources of financial support, with special mention of the late Martin Huberty, director of the University Water Resources Center.

I am also very grateful to Bruce Taylor and Martin Hackendorf for supervising the tritium measurements for several months, and to Joel Leventhal, Dan Benor, Arthur Gettleman, Gerry Marvray, and other undergraduate students for part-time assistance in making the measurements.

REFERENCES

- Begemann, E., and W. F. Libby, *Geochim. et Cosmochim. Acta*, **12**, 277, 1957.
- Brown, R. M., Hydrology of tritium in the Ottawa Valley, *Geochim. et Cosmochim. Acta*, 1960.
- Brown, R. M., and W. E. Grummitt, *Can. J. Chem.*, **34**, 220, 1956.
- Carlston, C. W., L. L. Thatcher, and E. C. Rodhamel, Tritium as a hydrologic tool—The Wharton tract study, Presented at IUGG meeting, Helsinki, 1960.
- Feely, H. W., and J. Spar, Tungsten 185 from nuclear bomb tests as a tracer for stratospheric meteorology, *Nature*, **188**, 1062, 1960.
- Gilletti, B. J., F. Bazan, and J. L. Kulp, *Trans. Am. Geophys. Union*, **39**, 807, 1958.
- Grosse, A. V., W. H. Johnston, R. L. Wolfgang, and W. F. Libby, *Science*, **113**, 1, 1951.
- Kaufman, W. J., and G. T. Orlob, *J. Am. Waterworks Assoc.*, **48**, 559, 1956.
- Kaufmann, S., and W. F. Libby, *Phys. Rev.*, **133**, 1954.
- Libby, W. F., Radioactive fallout particularly from the Russian October series, *Proc. Nat. Acad. Sci. U. S.*, **45**, 959, 1959.
- Libby, W. F., and C. E. Palmer, Stratospheric mixing from radioactive fallout, *J. Geophys. Research*, **65**, 3307, 1960.
- Lockhart, L. B., *USAEC Health and Safety Laboratory Quarterly Fallout Repts.*, 1960.
- Machta, L., Transport in the stratosphere and through the tropopause, *Advances in Geophysics*, **6**, 273, 1959.
- Martell, E. A., *Science*, **129**, 1197, 1959.
- Martell, E. A., and P. J. Drevinsky, *Science*, **1**, 1523, 1960.
- Stewart, N. G., R. G. D. Osmond, R. N. Crook, and E. M. Fisher, The world-wide deposition of long-lived fission products from nuclear explosions, *Atomic Energy Research Establishment (G. Brit.). Publ. AERE HP/R 2354*, 1960.
- Suess, H. E., Low level counting, *Interim Progress Rept., Contract AT(11-1)-34, Project 10*, August 1958a.
- Suess, H. E., Radioactivity of the atmosphere and hydrosphere, *Ann. Rev. Nuclear Sci.*, **8**, 2, 1958b.
- Von Buttlar, H., and W. F. Libby, *J. Inorg. Nuclear Chem.*, **1**, 75, 1955.
- Von Buttlar, H., and I. Wendt, *Trans. Am. Geophys. Union*, **39**, 660, 1958.

(Manuscript received May 26, 1961; revised July 31, 1961.)

Stability of Ice-Age Ice Sheets

J. WEERTMAN¹

*U. S. Army Cold Regions Research and Engineering Laboratory
Corps of Engineers, Hanover, New Hampshire*

Abstract. The stability of large ice sheets is investigated by using the present-day theory of the flow of ice in glaciers and ice sheets. The type of instability considered is that first mentioned by Bodvarsson. It is concluded that a small Arctic ice cap can become unstable and expand into a large ice age ice sheet as a result of moderate changes in the regime of the ice cap. A large continental ice sheet can also become unstable and shrink to nothing if the snow accumulation is reduced or the ablation rate increased. The results obtained fit well into the Milankovitch-Donn theory of ice ages. There is the possibility that the inherent instability of ice age ice sheets is in itself sufficient to explain both the formation and the disappearance of ice age ice sheets.

Introduction. In this paper we consider a problem important to any theory of the ice age. Namely, the stability of continental ice sheets. We wish to show from the mechanics of ice flow that a small ice cap situated in high latitudes on a continental land mass may become unstable in the sense that, if its width exceeds a critical size, the ice cap will grow unchecked as it reaches lower latitudes and is of continental dimensions. Further, it can be shown that once an ice cap reaches this size, another instability may set in if the rate of accumulation decreases or the rate of ablation increases. The ice sheet may then shrink to a small size or disappear.

We make the observation that an ice cap grows when the snowfall on it increases or melts at its edge decreases and that it shrinks when the snowfall decreases or the melt increases. This fact has been noted many times before. The new aspect introduced here is the use of the recently developed theory of the flow of glaciers and ice sheets to calculate the sensitivity of an ice cap to changes in rates of accumulation and ablation.

The type of instability with which we are concerned has already been pointed out by Bodvarsson [1955], who discovered this behavior while investigating ice sheet and glacier stability. His work was based on boundary con-

ditions rather different from those which would be used now.²

The instability discussed here is different in nature from that recently analyzed by Nye [1960]. Nye considered the problem of determining the manner in which a glacier or ice sheet approaches a stable steady-state profile. He showed that, in ablation regions, unstable behavior may occur before an equilibrium profile is finally reached. It is implicit in Nye's theory that there are always stable steady-state profiles which a glacier or ice sheet will approach. In other words, if the accumulation or ablation conditions change slightly, a glacier or ice sheet will assume a new steady-state shape of slightly different thickness and width. Nye was not primarily concerned with the problem that a steady-state profile might be in unstable equilibrium.

It may be useful to point out the reason why Nye's treatment leads to profiles of stable equilibrium whereas ours sometimes produces profiles that are unstable. Nye assumed that the accumulation area of a glacier or ice cap extends out to a fixed distance D from the center, as is shown in Figure 1a. Ablation occurs at all distances greater than D . Consider the successively larger ice caps labeled 1, 2, and 3 in Figure 1a.

² Bodvarsson used as the sliding velocity of an ice cap or glacier over its bed a quantity which is proportional to the stress acting at the bed and inversely proportional to the thickness of ice over the bed. In other words, he assumed $U \propto dh/dx$, whereas we shall use $U \propto h^2(dh/dx)^2$.

¹ Present address: Materials Science Department, Technological Institute, Northwestern University, Evanston, Illinois.

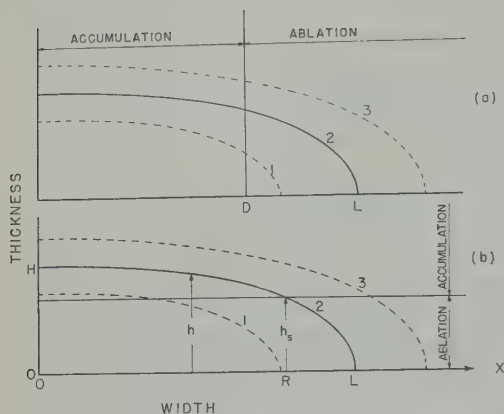


Fig. 1. Cross sections of one-half of ice sheets. (a) Accumulation area occurring out to a distance D from the center. (b) Accumulation area occurring above an elevation h_s .

It can be seen that regardless of the size of the ice cap the total accumulation area, and hence the total accumulation, remains constant, whereas the ablation area, and hence the total ablation, increases as the ice cap is made larger. Suppose curve 2 of Figure 1a represents an equilibrium ice cap, i.e., an ice cap whose total accumulation is balanced exactly by its total ablation. This ice cap obviously is in stable equilibrium. The total ablation of a larger ice cap (curve 3) would be greater than its total accumulation, a condition which would cause the ice cap to shrink. Similarly, the total ablation of a smaller ice cap (curve 1) would be less than the total accumulation, and the ice cap would grow towards the equilibrium size.

In contrast to Nye's assumption that the accumulation and ablation areas are determined by a horizontal distance D , our assumption is that the accumulation area is found above an elevation h_s (which will be considered later in the paper to be a function of the horizontal distance x) and the ablation area is found below this elevation, as is shown in Figure 1b. Under these conditions the accumulation area (and hence the total accumulation) is no longer constant but rather depends on the size of the ice cap. The larger the ice cap the larger the total accumulation. Suppose in Figure 1b that curve 2 represents an equilibrium profile. It is now open to question whether this profile is of stable equilibrium. The total accumulation of a larger ice cap, represented by curve 3 in Figure

1b, could be greater than its total ablation, in which case the ice cap would grow rather than shrink toward the equilibrium size. Similarly, a smaller ice cap, curve 1, could have a total ablation greater than its total accumulation, and hence it would shrink and ultimately disappear.

Preliminary considerations. There are three main variables which are involved in the termination of the size of an ice cap or ice sheet. These are the accumulation rate in the accumulation area which exists at the high elevations of an ice cap (see Fig. 1b), the ablation rate in the ablation area of the lower elevations, and the elevation of the border separating the accumulation area from the ablation area. To simplify the analysis it will be assumed that the accumulation rate at every point in the accumulation area is equal to the average accumulation rate a and that everywhere in the ablation area the ablation rate is the average ablation rate \bar{a} . This assumption does not involve too great a loss in the accuracy of the results since it is known that accumulation and ablation rates usually are slowly varying functions of the distance x from the center of an ice sheet. Thus the average ablation or accumulation rate is almost always of the same order of magnitude as the actual accumulation or ablation rate.

One can expect that the average accumulation and ablation rates will be a function of the size of the ice cap or ice sheet. Conversely, one can also expect that the size of an ice cap depends on the magnitude of its accumulation and ablation rates. To disentangle these two functional relationships we shall assume in this analysis that the accumulation rate and ablation rate are constant and independent of the size of the ice sheet. The effect of a change in sheet size on the accumulation and ablation rates can be examined later by means of the equations obtained upon the assumption of constant rates.

Once the average accumulation and ablation rates are specified, an ice sheet profile can be calculated from the principle of conservation of mass [Nye, 1959]. Only two-dimensional sheets (i.e., ice sheets which have one axis much longer than the other and whose profile, therefore, are essentially independent of distance along the long axis) will be considered in this paper. As Nye has shown, the profile of a circular ice sheet is essentially the same as that

-dimensional ice sheet, and thus the obtained from two-dimensional ice sheets apply well to circular ice sheets.

sheet profile is calculated [Nye, 1959] finding the volume of ice passing through a section of the ice sheet to the total of ice accumulation between that cross and the center of the ice sheet. If the accumulation and ablation rates are equal and if the elevation of the border separating accumulation and the ablation areas is specified, the analysis of the following shows that *only one possible equilibrium profile exists*.

Line of fixed elevation. We define an elevation h_* (called hereinafter the elevation of the line) to be an elevation such that any part of an ice sheet which lies above h_* is in an accumulation zone and any part which lies below h_* is in an ablation zone. The base of the ice sheet is assumed to be flat and at sea level. In this section it is assumed that the value of U is constant.

U represents the average velocity of ice passing through a cross section situated at a distance x from the center of the ice sheet. The total volume of ice passing through this section per unit time (for a unit length of ice sheet) is Uh , where h is the height of the ice sheet above the bed. Let A represent the accumulation or ablation at any point on the ice sheet surface. For steady-state conditions one has the equation

$$Uh = \int_0^x A \, dx \quad (1)$$

It is assumed that the average velocity U is constant, and the equation

$$U = B\tau^m \quad (2)$$

is a constant, m is a constant whose value is of the order of 2 to 2.5, and τ is the shear stress acting at the bed of an ice sheet. The value of τ is given by

$$\tau = -\rho gh(dh/dx) \quad (3)$$

In Nye's analysis it is assumed that the shear stress is the dominant stress and that longitudinal stresses are unimportant. It is possible [Weertman, 1957] to calculate ice sheet profiles when longitudinal stresses are large. Since the profiles so obtained are almost the same as those found by Nye, we do not need to use this more complicated theory in the above analysis.

where ρ is the density of ice, g is the gravitational acceleration, and dh/dx is the slope of the upper ice surface. It is reasonable to use the velocity given by (2), since a velocity of this form is to be expected both from an analysis of the differential flow of ice within an ice mass [Nye, 1959] and from a theoretical study of the sliding of ice [Weertman, 1957] over a glacier or ice-cap bed. Since the value of m in (2) is approximately 2, we shall assign this value to it now in order to avoid some cumbersome expressions in later sections of this paper. When (2) and (3) are substituted into (1) and m is set equal to 2, one obtains

$$B(\rho g)^2 h^3 (dh/dx)^2 = \int_0^x A \, dx \quad (4)$$

The solution of this equation is the profile of an equilibrium ice sheet.

The average accumulation rate a and the average ablation rate \bar{a} are defined by

$$a = R^{-1} \int_0^R A \, dx \quad (5)$$

$$\bar{a} = -(L - R)^{-1} \int_R^L A \, dx$$

where R is the distance from the ice sheet center to the border between the accumulation and ablation areas and L is the width of the ice sheet (Fig. 1). The term \bar{a} is defined so that it is a positive quantity. If equations 5 are substituted into (4) and if the requirement for an ice sheet to be in equilibrium,

$$aR = \bar{a}(L - R) \quad (6)$$

is satisfied, the following equations are obtained for the equilibrium profile:

$$h^{5/2} = H^{5/2} - (c/\bar{a})(ax)^{3/2} \quad (7)$$

(when $h > h_*$ and $0 \leq x \leq R$)

$$h^{5/2} = h_*^{5/2} + (c/\bar{a}) \cdot \{[aR - \bar{a}(x - R)]^{3/2} - [aR]^{3/2}\} \quad (8)$$

(when $h < h_*$ and $R \leq x \leq L$)

where H is the thickness of the ice cap or ice sheet at its center and c is equal to $(5/3)/B^{1/2}\rho g$. The distance R is given by

$$R = a^{-1}(\bar{a}/c)^{2/3} h_*^{5/3} \quad (9)$$

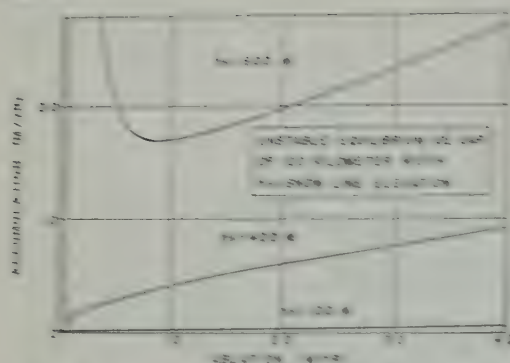


Fig. 2. Curves of accumulation versus distance for an unstable ice sheet of 100-km width for three different snow line elevations.

the thickness H by

$$H = (1 + a/c)^{1/2} L \quad (10)$$

and the distance L from center to edge by

$$\begin{aligned} L &= [(1 + a/c)^{1/2} R^2 - L_0^2]^{1/2} \\ &= R^2 [(1 + a/c)^{1/2} - L_0^2/R^2]^{1/2} \\ &= [1 + a/c]^{1/2} R \end{aligned} \quad (11)$$

The profile determined by (7) through (11) is one of unstable equilibrium. For example, both the width L and the thickness H decrease with increasing accumulation rate or lowering snow line. Now an increase in the accumulation rate or a lowering of the snow line obviously increases the volume of an ice sheet. However, since the new equilibrium profile determined by (7) through (11), resulting from this change has a smaller total volume, a new equilibrium profile will not be approached if the accumulation rate is increased or the snow line lowered. The ice sheet will grow indefinitely. Similarly, if the accumulation rate is decreased or an ice sheet in equilibrium or the snow line is raised, an ice sheet or ice cap will shrink until it disappears, since the new equilibrium profile contains a greater volume of ice than the original equilibrium profile.

In Figures 2 and 3 are shown graphs of the values of accumulation versus distance which satisfy (7) through (11) for various values of the elevation of the snow line and when the width of the ice cap is either 100 or 3000 km (these values are the total width = $2L$). A value of $c = 2$ (meter-years)^{1/2} was used in

making the calculations. The value of c is obtained from a comparison of theoretical snow profiles with measured profiles (Weertman, 1964a). Values of accumulation and ablation rates which lie above a given curve in Figs. 2 and 3 would lead to an ice sheet, not originally 100 or 3000 km wide, growing indefinitely in size. If the accumulation or ablation rates are below the curve, the ice cap or sheet will shrink away to nothing. From Figure 2 it can be seen that rather small rate of accumulation will lead to the unstable growth of an ice cap 100 km wide, if the snow line is as low as 100 m. On the other hand, if the snow line is at 300 m, a very high rate of accumulation (~ 2 m/yr) will be required to maintain the ice cap or cause it to grow.

Stability and instability considerations. In the previous section it was shown that, if accumulation and ablation rates are fixed, and the snow line elevation is fixed, only one possible equilibrium ice sheet profile exists and this profile is of unstable equilibrium. Stability is introduced only if a , L , or L_0 is some appropriate function of ice sheet size. For example, suppose the snow line elevation is given by

$$L_0 = a^2 c^2 / R^2 = R^2 c^2 / a^2$$

This equation is simply (6) rearranged to have located the snow line elevation of equilibrium L_0 in order to emphasize the fact that it is a snow line elevation which obeys a particular equation. We shall call this particular snow line elevation the elevation of the equilibrium firm line. When the snow line deviates

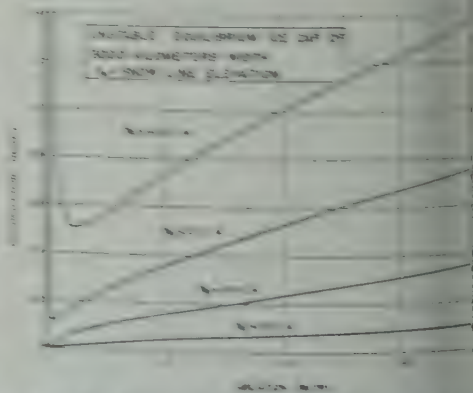


Fig. 3. Curves of accumulation versus distance for an unstable ice sheet of 3000-km width for four different snow line elevations.

the ice velocity of accumulation, and the ice sheet will retreat to a new equilibrium position if, initially, it is either too small or too large.

The lower line will not follow an h of (12). The dotted curve line is an approximately linear one, a function of h (cf. Hays and Rind, 1975). The curve for h_c is shown in Figure 4. Schematically in this figure is h_c with \dot{a} and \dot{a} held constant. An ice sheet can retreat only when there is a net loss of ice, that is, when $h_c > h$. The profile of position of the ice sheet, if they are limited to each side, is possible. If they were each shown in the figure, two profiles are

one of the stability of an ice sheet. A constant in the crossing of the two lines h_c can be reached by a constant value profile, which is slightly higher than the equilibrium position. In the case of a somewhat low the distance R due to the border of accumulation and ablation areas is small, the corresponding velocity difference in sheet. For the larger the distance R and h_c in Figure 4 equal. If h_c is less than h , the larger the sheet will grow and if h_c is greater the sheet will shrink. In order for the ice sheet to be stable it must not be too large or too small. It is necessary larger than the equilibrium and to grow if it becomes smaller than h_c . If h_c is variable the opposite. Thus the left-hand intersection of h_c and h can be variable equal to the right-hand intersection. The ice sheet will not a value less than that of the left-hand intersection and it disappears. If h_c is larger than that of the left-hand h_c , the sheet will grow and it reaches a new profile represented by the right-hand intersection.

As in that of Figure 4 can be re-considering h_c to be fixed and assuming \dot{a} to be a function of the sheet size. It is possible to show the accumulation of the sheet would have to decrease as the rate increases with increasing size

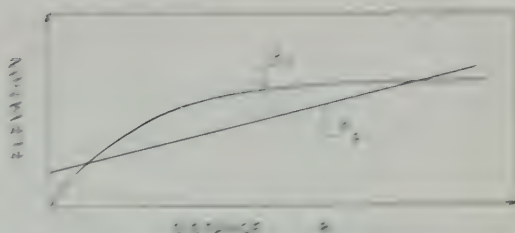


Fig. 4. Schematic plot of the elevation of the ice sheet h , and the elevation of the equilibrium line h_c of ice accumulation for one same distance R from the center of an ice sheet to the boundary between the accumulation and ablation areas.

of the ice sheet. Hence it is virtually impossible to discuss the functional relationship between accumulation and ablation rates and the size of an ice-age ice sheet, we shall consider the influence of accumulation and ablation only through the equations obtained when h_c is allowed to vary.

Stability of Antarctica and Greenland ice sheets. Before proceeding any further we should like to make a statement in order to point out the fact that the ice sheet of Antarctica (and to a lesser extent, Greenland) is a rather impervious case in the analysis. We are assuming that we are dealing with an ice sheet which lies on a land base extending to infinity. In Antarctica, and to a lesser extent, in Greenland, the ice sheet terminates at the water's edge where the ice can be broken, providing for an infinite amount of ablation. This case if h_c is at sea level and there is no ablation area on the ice sheet itself, a stable ice sheet can exist because there is no more land on which to intensify accumulation ice sheet can spread. The profile of such an ice sheet can be obtained from (7) through (11) by allowing the thickness to be very large at the edge of the ice sheet and from (11) by setting $\bar{F}^{2L,50} = Lc^{50}$, where $2L$ is the width of the land mass. In the equation $R \approx L$ and the ablation starts virtually at the edge of the ice sheet.

Profile of an ice sheet with constant accumulation and ablation rates and a variable snow line. In this section we calculate the profile of an ice sheet which has a fixed rate of accumulation and ablation, but whose snow line is a function of distance. We know of course, that the snow line is high located but at low elevations

and that it rises to elevations of the order of 4 km in temperate latitudes [Matthes, 1942; Charlesworth, 1957]. The rise appears to be linear [Fig. 6 on p. 9 of Charlesworth, 1957]. Thus we can write for h .

$$h_s = \bar{h} + sx \quad (13)$$

where x is measured toward the equator, s is a constant of the order of 10^{-3} , and \bar{h} is the value of h_s in polar regions (at $x = 0$). Its value [Matthes, 1942; Charlesworth, 1957] seems to lie in the range from 100 to 1000 meters.

Suppose we consider an ice sheet whose center lies at $x = 0$, and the snow line rises in either direction from the center of the sheet by an equation of the form of (13). The analysis of the section of a snow line of fixed elevation is still valid, with the exception of (9), which, in order to take (13) into account, should be replaced by

$$h_s = \bar{h} + sR = a^{3/5}(c/\bar{a})^{2/5}R^{3/5} = h_r \quad (14)$$

There may be two values, one value, or no value of R which satisfies this equation. These values of R represent the points of intersection of the curves h_r and h_s shown in Figure 4. If the two curves of that figure never meet (this occurs when a is small, or \bar{a} or s is very large) there is no possible equilibrium ice sheet, and any ice sheet already in existence will shrink and disappear. If s is equal to zero, only one intersection point exists. This is the case considered in a previous section, and there we found that the ice sheet is not stable. If s is increased from zero to some finite value, two intersections will occur. The intersection with the smaller value of R gives a profile that approximates the case in which s is equal to zero and clearly corresponds to an unstable ice cap. Consider the intersection at the larger value of R . Let us take the case in which R is so large that $sR > \bar{h}$. In this situation, R is approximately equal to

$$R = (1/s)^{5/2}(c/\bar{a})a^{3/2} \quad (15)$$

and h_s , H , and L become

$$h_s = (1/s)^{3/2}(c/\bar{a})a^{3/2} \quad (16)$$

$$H = (1/s)^{3/2}(1 + a/\bar{a})^{2/5}(a/\bar{a})^{3/5}a^{1/2}c \quad (17)$$

$$L = (1/s)^{5/2}(1 + a/\bar{a})(c/\bar{a})a^{3/2} \quad (18)$$

The magnitudes of H and L in these equations increase with increasing accumulation rate and

decreasing ablation rate and decrease with increasing a and increasing \bar{a} . Hence, the profile can be a stable one. If the rates of accumulation and ablation are changed, the ice sheet will be able to approach a new equilibrium profile.

Effect of change in position of center of ice sheet. The center of an ice sheet located on a continent in the northern hemisphere will migrate southward as the ice sheet grows. Because of this migration of the center, the analysis of the previous section has to be modified slightly. The height of the equilibrium firn line still is

$$h_r = a^{3/5}(c/\bar{a})^{2/5}R^{3/5}$$

where R is measured from the center, but the snow line elevation at R is now

$$h_s = \bar{h} + s(L' + R)$$

where L' is the width of the northern half of the ice sheet. If the rate of accumulation is the same in the northern and southern halves of the ice sheet then according to (11) the widths of the two halves will be approximately the same.⁴

If h_s is set equal to h_r and $L = L' \cong R$ we obtain

$$\bar{h} + 2sR = a^{3/5}(c/\bar{a})^{2/5}R^{3/5} \quad (19)$$

which replaces (14) of the previous section. Using this equation we find that

$$R = (1/2s)^{5/2}(c/\bar{a})a^{3/2}$$

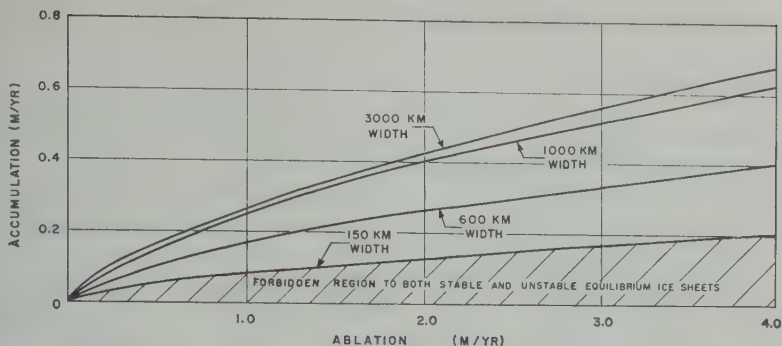
$$h_s = (1/2s)^{3/2}(c/\bar{a})a^{3/2}$$

$$H = (1/2s)^{3/2}(1 + a/\bar{a})^{2/5}(a/\bar{a})^{3/5}a^{1/2}c$$

$$L = (1/2s)^{5/2}(1 + a/\bar{a})(c/\bar{a})a^{3/2}$$

⁴ By center of an ice sheet we mean the position of greatest elevation where $h = H$. According to (11) whenever \bar{a} is larger than a (the commonly occurring relationship) then $L/(a^2)^{-1/5}H^{5/5}$ and $L'/(a^*)^{-1/5}H^{5/5}$, where a^* is the accumulation rate in the northern half, must be the same for the northern and southern halves. We see that $L/L' \cong (a^*/a)^{1/5}$. Since ratio L/L' depends only on a one-third power a^*/a the widths of the southern and northern halves will almost equal each other if a^* is of the same order of magnitude as a . The analysis is easily modified to account for the case in which L differs greatly from L' . When L approaches L' it is also almost equal to R .

The northern half of the ice sheet is stable, of course, in the sense that the Antarctic ice sheet is stable, since it cannot spread out indefinitely in a northerly direction.



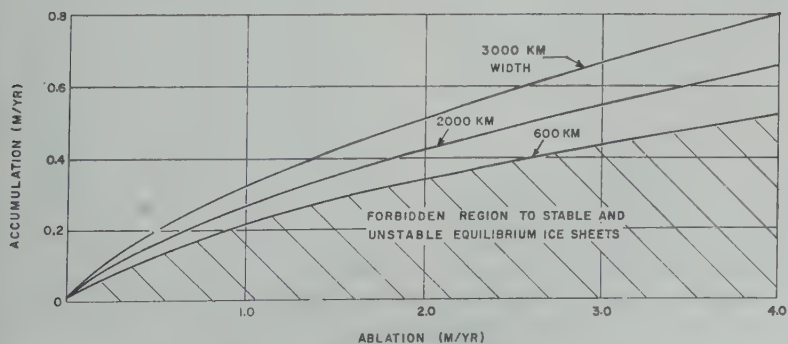
g. 5. Curves of accumulation versus ablation for stable equilibrium ice sheets of four different widths when the snow line elevation is a linear function of distance ($\bar{h} = 100$ meters, $s = 10^{-3}$).

How sensitive the width of an ice sheet in stable equilibrium is to accumulation conditions, we have plotted curves of accumulation versus ablation for ice sheets of different widths (Figs. 5 and 6). We have set s at chosen values for \bar{h} of 100 meters and 400 meters. The width in these figures is defined to be twice R , and R is determined from the condition that for a fixed value of \bar{h} , there is a minimum ablation below which it is impossible for an ice sheet or ice cap to exist in either stable or unstable equilibrium. (This minimum width occurs at the point where the two curves of Figure 4 are tangent to one another.) The region shown in Figure 5 is the forbidden region where equilibrium ice sheets or ice caps cannot exist. This region is labeled the forbidden region.

It should be noted in these figures that small changes in the rate of accumulation can change the width of an ice sheet in stable equilibrium by very large amounts and can even make

it impossible for such an ice sheet to exist. The width of a stable ice sheet is particularly sensitive to s , the rate of change of the elevation of the snow line. To illustrate this sensitivity, we have plotted in Figure 7 curves of accumulation versus width of ice sheets in equilibrium (both stable and unstable), for a fixed value of ablation. Two different values of s were used when calculating these curves. On each curve, points to the right of the minimum represent stable equilibrium widths and points to the left unstable widths. The curve corresponding to the smaller value of s has a smaller slope in the stable region. Hence, changing the rate of accumulation results in a larger effect on the width of the ice sheet in equilibrium. The smaller s is, the more sensitive the width of a stable ice sheet is to changes in accumulation (and also in ablation).

It is known [Charlesworth, 1957] that during the ice age the snow line was lowered approxi-



g. 6. Curves of accumulation versus ablation for stable equilibrium ice sheets of three different widths when the snow line elevation is a linear function of distance ($\bar{h} = 400$ meters, $s = 10^{-3}$).

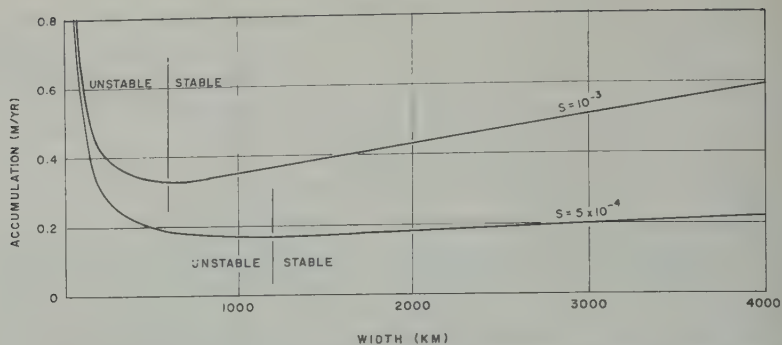


Fig. 7. Curves of accumulation versus equilibrium width of ice sheets when the snow line is a linear function of distance, the ablation rate is 2 meters/year, and \bar{h} is 400 meters.

mately 1000 meters in temperate latitudes. Thus when an ice-age ice sheet reaches a large size the equation we have used for h_s can probably still be employed but with a value of s which is 30 to 50 per cent smaller than its present-day value.

Application to glaciers: The analysis we have made of the stability of an ice sheet resting on a flat base can also be applied to a glacier descending a mountain or to a small ice cap which has formed in an elevated region and whose edge must descend into a lower elevation as it expands.

A glacier normally rests on a sloping bed. If the slope of the bed is β_s the equation which describes the profile of the glacier is

$$B(\rho g)^2 h^3 (\beta + dh/dx)^2 = \int_0^x A dx \quad (25)$$

In this equation x is measured parallel to the bed. A profile similar to that of an ice sheet will be obtained from this equation.

Since the coordinate system for the glacier is tilted at an angle β compared with a system whose horizontal distance is parallel to the earth's surface, the snow line elevation will be given by

$$h_s = \bar{h} - \beta x \quad (26)$$

where s has been set equal to $-\beta$ since β is less than zero. We have assumed in (26) that the elevation of the snow line in the mountains remains at a constant elevation above sea level. The value of \bar{h} depends on the location of the origin of x . Typical values of β range around 10^{-2} radians, an order of magnitude larger than what we have used for s . From (18), which should be qualitatively correct when applied to

glaciers, it can be seen that the length of the stable equilibrium profile depends inversely on s to a $5/2$ power, if $sR > \bar{h}$. Stable equilibrium profiles for glaciers should occur at lengths which are of the order of 100 to 1000 times smaller than those of continental ice sheets. This situation results from the fact that $-\beta$ is at least 10 times larger than the values of s appropriate to large ice sheets. Since continental ice sheets had dimensions of the order of 2000 km, stable equilibrium glaciers could be expected, and do, exist for lengths of the order of 1 to 10 km. The fact that the average ablation rate increases the further a glacier descends down a mountain also stabilizes a glacier. However, this increase is not essential to insure stabilization.

Discussion. The analysis presented in this paper suggests that a small ice cap can become unstable and grow to a large size if it exceeds a critical width. The critical nucleation size could be of the order of the dimensions of existing ice caps in the Arctic (~ 30 km) when the snow line elevation is of the order of 100 to 400 meters and the product (ablation rate) $^{2/3}$ /(accumulation rate) is less than, or equal to, 1 to $(\text{m/yr})^{-1/3}$ (for example, an ablation rate of 0.1 m/yr and an accumulation rate of 1 m/yr). If an ice cap is less than this critical size, it should disappear completely. The fact that existing ice caps persist can be explained by the reasons mentioned in the previous section or by local weather peculiarities.

In theories of the ice ages⁵ it is usually assumed that an ice age starts as a result of

⁵ Reviews of the more important work in this field are given by Flint [1947] and Charlesworth [1957].

accumulation or decreased ablation rates, the conditions which would be required to make a small ice cap 'go critical.' Ewing and Donn [1956, 1958], for example, proposed that an ice age begins when the Arctic Ocean becomes ice free, a situation which leads to increased snowfall on the lands surrounding this ice-free water. It may not be necessary, however, to invoke some special event to make a small ice cap go critical. Normal weather fluctuations may produce a century of greater than normal accumulation and less than normal ablation; these conditions then may induce a small ice cap to start growing to a large size.

Even if a small ice cap did grow into a large ice-age ice sheet, is it possible for the sheet ever to melt away again without a significant change in the weather conditions (other than that produced by the ice sheet itself)? The question on the ice sheet itself would be whether it is likely to decrease as it became bigger, both in terms of the accumulation area would be at a maximum and because the cooling of the ice sheet in the presence of the sheet might lead to increased precipitation. The Antarctic ice sheet has rates of accumulation (~ 10 to 20 cm/year), as does the smaller Greenland ice sheet (~ 10 to 40 cm of ice/year), and a large ice sheet might also be expected to have similar values. An inspection of Figures 2 to 7 shows that when accumulation rates fall to low values, an ice cap may easily become unstable and shrink to nothing. If low accumulation rates once set in, and if they persist, there is no serious problem in explaining the existence of ice-age ice sheets. However, one might expect that as a large ice cap began to melt, the accumulation rate would increase and the ablation rate decrease as the edge retreated (poleward). Ewing and Donn circumvented this difficulty by their proposal that the Arctic Ocean, which was ice free at the beginning of an ice age, freezes over again and the ice cover remains until after the complete disappearance of the continental ice sheet. According to their theory, the snow precipitation rates are low over the Arctic Ocean, and low rates occur over the Arctic Ocean is covered with ice. If this theory is correct, there is no difficulty in understanding why an ice-age ice sheet suddenly becomes unstable and shrinks to nothing. If the ice cap grows and the Arctic Ocean

freezes over, the accumulation rate decreases until the instability discussed in this paper overtakes the ice sheet, which then shrinks until it disappears.⁶

Our results answer one major criticism [Livingstone, 1959] of the Ewing-Donn theory. This criticism is that the theory contained no mathematically demonstrated instability which could lead to ice ages. It partially answers another serious point raised against their theory [Livingstone, 1959], namely, that the existence of unglaciated regions around the Arctic Ocean, such as Peary Land, argues against the theory. This latter criticism assumes tacitly that an ice age starts in the Ewing-Donn theory with *such a very heavy* rate of snowfall on the lands around the Arctic Ocean that no land can remain ice free. Now if only a modest increase in the rate of snowfall (say by a factor of 2 or 3) is required to make a small ice cap go critical, there is no reason to expect that all the lands around the Arctic Ocean will become ice covered even though an ice age can occur. Thus the existence of unglaciated regions in the Arctic is not a fatal objection to the Ewing-Donn theory.⁷

There is another way in which an instability in an ice-age ice sheet might be brought about. A small ice cap starting to grow in the far north is likely to be frozen to its bed. If such is the case, the ice cap will not be able to slide over its bed, and the effective value of B in (2) will be reduced and the value of c increased. As the ice sheet grows, it may still remain frozen to its bed until it reaches a large size. Small thickness and large accumulation rates favor a cold ice cap being frozen to its base [Robin, 1955; Weertman 1961b], and large thickness and small accumulation rates increase the probability that the bottom will be at the melting point. Now suppose that when the ice sheet reaches a large size, the temperature at bottom rises until the ice there is at the melting point, and thus the ice can slide over its bed. The value of c will be de-

⁶ The ice-age ice sheets did not retreat in a continuous manner but rather the recession was interrupted time and again by readvances of the ice edge. This fluctuation in the dimensions of ice-age ice sheets very probably can be explained through Nye's theory [Nye, 1960] of the inherent instability of the ablation region of glaciers and ice sheets.

⁷ Ewing and Donn [1959] have given other replies to these two criticisms.

creased. From studies made on temperate glaciers of the fraction of motion which is due to sliding and the fraction which is due to differential motion within the glaciers, one can estimate that c will be decreased by a factor of about $\frac{1}{2}$. An inspection of the equations developed in this paper will show that a decrease in c of $\frac{1}{2}$ is essentially equivalent to an increase in the ablation rate of a factor of 2. Such an increase in the ablation rate might lead to the instability of an ice-age ice sheet and its subsequent shrinkage. The decreased value of c would persist as the ice sheet decreased in size for the following reason. The bottom of a cold ice cap can be at the melting point because of the geothermal heat flowing up through the earth. In addition, if an ice cap is sliding on its bottom, the heat of sliding helps to keep the bottom at the melting point. The heat of sliding is of the same order of magnitude as the geothermal heat [Weertman, 1957, 1961b]. Because of this extra heat from the sliding process itself, once an ice cap originally frozen to its bed starts to slide it could persist in sliding even though brought back to the conditions where formerly it had been frozen.⁸

Another factor which may be responsible for initiating instability in a continental ice sheet is the isostatic sinking of the ice sheet bed as the ice sheet becomes large. The ice sheet profile can be found quite easily when this sinking occurs [Weertman, 1961a]. In such a situation the equation determining the profile is

$$B(\rho g)^2 \gamma^3 h^3 (dh/dx)^2 = \int_0^x A dx \quad (27)$$

where h is now the present height of the upper ice sheet above the original position of the bed before sinking occurred and $\gamma = (1 - \rho/\rho_r)^{-1}$ where ρ_r is the average density of rock below the ice sheet ($\rho_r \approx 3\rho$). Equation 27 can be considered to be the same as (4) but with a value of B which is $\gamma^3 \approx 3.4$ times larger. As before, this apparent increase in B can be considered to be equivalent to an increase in the ablation rate by a factor of about 1.8. Such an increase could possibly set off an instability in an ice sheet. If this mechanism is to be effective, however, the time for isostatic sinking to occur must

be longer than the time required to build up a large ice sheet.

Summary. By utilizing the present-day theory of the flow of ice in glaciers and ice sheets we have shown that a small Arctic ice cap can expand into an ice-age sheet without large changes in the rate of snow accumulation or ablation, or in the elevation of the snow line. Once such an ice cap has expanded, it can become unstable again and shrink to nothing if the accumulation rate is reduced or the ablation rate increased and if the reduction or increase persists as the ice sheet shrinks. These results are in good agreement with the recent Ewing-Donn theory of ice ages. There is a possibility that the inherent stability of ice-age ice sheets is in itself sufficient cause to account for their expansion and subsequent shrinkage.

Acknowledgments. I appreciate the critical comments of Prof. H. Bader, Prof. R. Sharp, and J. F. Nye on an earlier version of this paper. I have no responsibility, of course, for any possible erroneous results or conclusions in the present version.

REFERENCES

- Bodvarsson, G., On the flow of ice-sheets and glaciers, *Jökull*, 5, 1-8, 1955.
- Charlesworth, J. K., *The Quaternary Era*, Edward Arnold Ltd., London, vol. 1, p. 9, and vol. 2, p. 652, 1957.
- Ewing, M., and W. L. Donn, A theory of the ice ages, *Science*, 123, 1061-1066, 1956; *ibid.*, 129, 1159-1162, 1958; *ibid.*, 129, 464-465, 1959.
- Flint, R. F., *Glacial Geology and the Pleistocene Epoch*, John Wiley & Sons, New York, 1947.
- Livingstone, D. A., Theory of ice ages, *Science*, 129, 463-464, 1959.
- Matthes, F. E., *Glaciers, Hydrology (Physics of the Earth, 9)*, McGraw-Hill Book Co., New York, 1942.
- Nye, J. F., The motion of ice-sheets and glaciers, *J. Glaciol.*, 3, 493-507, 1959.
- Nye, J. F., The response of glaciers and ice-sheets to seasonal and climatic changes, *Proc. Roy. Soc., London* 256A, 559-584, 1960.
- Robin, G. de Q., Ice movement and temperature distribution in glaciers and ice-sheets, *J. Glaciol.*, 2, 523-532, 1955.
- Weertman, J., On the sliding of glaciers, *J. Glaciol.*, 3, 33-38, 1957.
- Weertman, J., Equilibrium profile of ice caps and ice sheets, *J. Glaciol.*, (in press) Oct. 1961.
- Weertman, J., Mechanism for the formation of inner moraines found near the edge of cold ice caps and ice sheets, *J. Glaciol.*, (in press) 1961b.

⁸ Robin [1955] used the arguments given in this paragraph to explain catastrophic advances of glaciers.

Depth to Sources of Magnetic Anomalies

LEROY R. ALLDREDGE AND GERALD D. VAN VOORHIS

*U. S. Coast and Geodetic Survey
Washington, D. C.*

Abstract. The characteristics of several long magnetic total field intensity profiles have been determined. Only track lines which were nearly straight and longer than 2000 miles were considered. First, the centered dipole field total intensity was subtracted from the measured intensity to obtain a real nondipole field. A smooth curve was then drawn through this dipole field using a stiff spline curve. The distance between crossover points of the smooth and the nondipole field was determined and a histogram of the results plotted. The results confirm the earlier results of Serson and Hannaford that most of the anomalies have a very short wavelength. Ninety-three per cent of the cases had crossover distances less than 60 nautical miles. The simple form of the smooth curves indicated a nearly sinusoidal departure from a dipole field having crossover points between 2100 and 5200 nautical miles. The most natural unforced explanation of the above results is that short-wavelength anomalies are due to crustal effects and the long-wavelength anomalies are due to causes within the interior of the earth. The large gap between the short- and long-wavelength groupings supports the hypothesis that the mantle is a forbidden region for magnetic sources. This conclusion is illustrated by calculations based on simple models.

Introduction. Magnetic anomalies no more than a few tens of miles in width (short-wavelength anomalies) are used to infer the underground structure in prospecting. World charts, on the other hand, do not preserve anomalies less than a few hundreds of miles in width (long-wavelength anomalies). In this study the intermediate-wavelength region is investigated to determine if there is a natural division of anomalies into short- and long-wavelength categories.

Very recently, small-scale world-wide magnetic charts were based entirely on discrete measurements at observatories, repeat stations, and other individual points throughout the world. It has been the usual practice in compiling world-wide charts to assign to each 1° square the mean of all reliable measurements made within the quadrangle. In areas where measurements are sparse, and especially in polar regions, larger quadrangles are used. It is obvious that such a procedure tends to obscure anomalies less than 50 to 100 miles across. The resulting data are contoured using a practice which introduces further smoothing. The resulting charts contain none of the gradients caused by so-called local anomalies which may on occasion exceed 1000 gamma in amplitude and may extend over distances from 150 to 200 miles.

Because the charts have been derived from discrete measurements, spaced at varying distances, there has always been some question about the nature of the anomalies which were discarded in the smoothing process. The charts undoubtedly describe the broad regional magnetic effects quite faithfully, but they leave uncertainty about the magnetic structure on a finer scale.

At the other end of the charting spectrum we have detailed charts of total field intensity constructed from continuous profile records made with an aircraft flying at a low altitude along closely spaced lines [Henderson and Zietz, 1958]. These charts normally cover a limited area of several hundred miles square. They give the detailed magnetic structure but, because of their limited coverage, often fail to indicate broad-scale regional trends. Very often in the analysis of such records an attempt is made to subtract the broad regional magnetic trends so as to highlight the local anomalies.

Knapp [1942], in the construction of the 1940 isogonic chart of the United States, clearly recognized the difference between local anomalies and a datum chart containing the earth's main field and including broad regional anomalies. The United States isogonic chart is based on nearly 10,000 individual values and hence contains more detail than the world charts.

Knapp suggested that isanomalic charts which would be fixed in amplitude and position be drawn up to represent the local anomalies and that these be added to a newly determined datum chart for each new epoch. Knapp noted that constructing charts by this method tended to show the absence of anomalies having a scale intermediate between that of the fine structure of the local anomalies and that of the general distribution reflected in the datum chart.

Vestine, Laporte, Lange, and Scott [1947] indicated that in drawing isomagnetic charts over ocean areas there was the suggestion of possible anomalies 1000 to 2000 km in linear cross section, but that they may be only apparent rather than real, and that they should be checked by means of total intensity measurements made from aircraft.

In this paper several records of total field intensity are examined. They were taken along track lines which run several thousand miles in a nearly straight line. This gives a real opportunity to study not only the local anomalies and the broad regional anomalies but also the

region in between. *Serson and Hannay* [1957] have indicated that in western Caribbean profiles of the magnetic components resemble gentle curves with wavelengths of a few thousand kilometers on which are superimposed regular fluctuations. In their investigation the shorter-wavelength irregular fluctuations they first subtracted the long-wavelength gentle curves. In this study all wavelengths are retained and analyzed in a simple statistical manner.

Secular variations have no bearing on the arguments used here because of their extremely slow rates of change. More rapid variations corresponding to magnetic storms and diurnal variations do add undesirable noise to the analysis but there was no method available for removing this noise. However, this limitation is not considered serious because the amplitudes are generally smaller than the corresponding amplitudes of the areal anomalies being studied.

Descriptive and statistical analysis. A series was made for very long magnetic total intensity profiles taken along track lines appearing

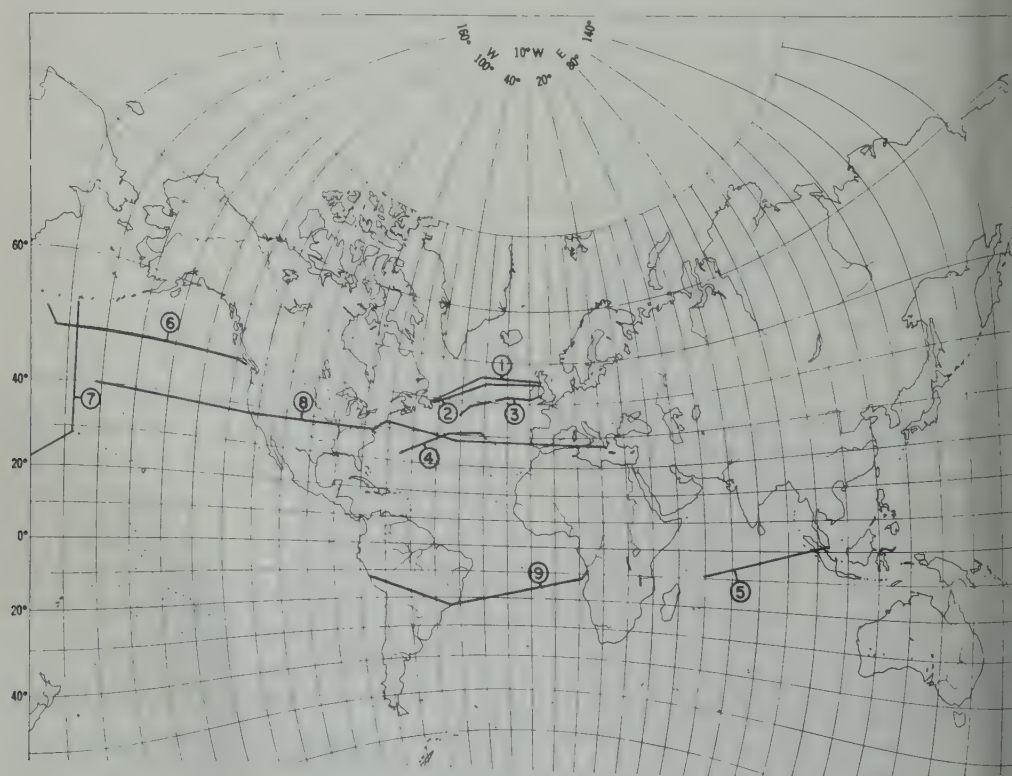


Fig. 1. Index map showing the general location of track lines.

geodesics. It was considered necessary that track lines be at least 2000 miles long and not contain large sudden changes in direction. Shorter track lines would give information about the short-wavelength features but not yield information about the longer-

wavelength characteristics. Nearly all of the aeromagnetic surveys for geophysical prospecting are made up of flight lines containing straight segments only a few hundred miles in length and therefore not usable.

Track lines 1, 2, 3, 4, 5, 8, and 9 of Figures 1,

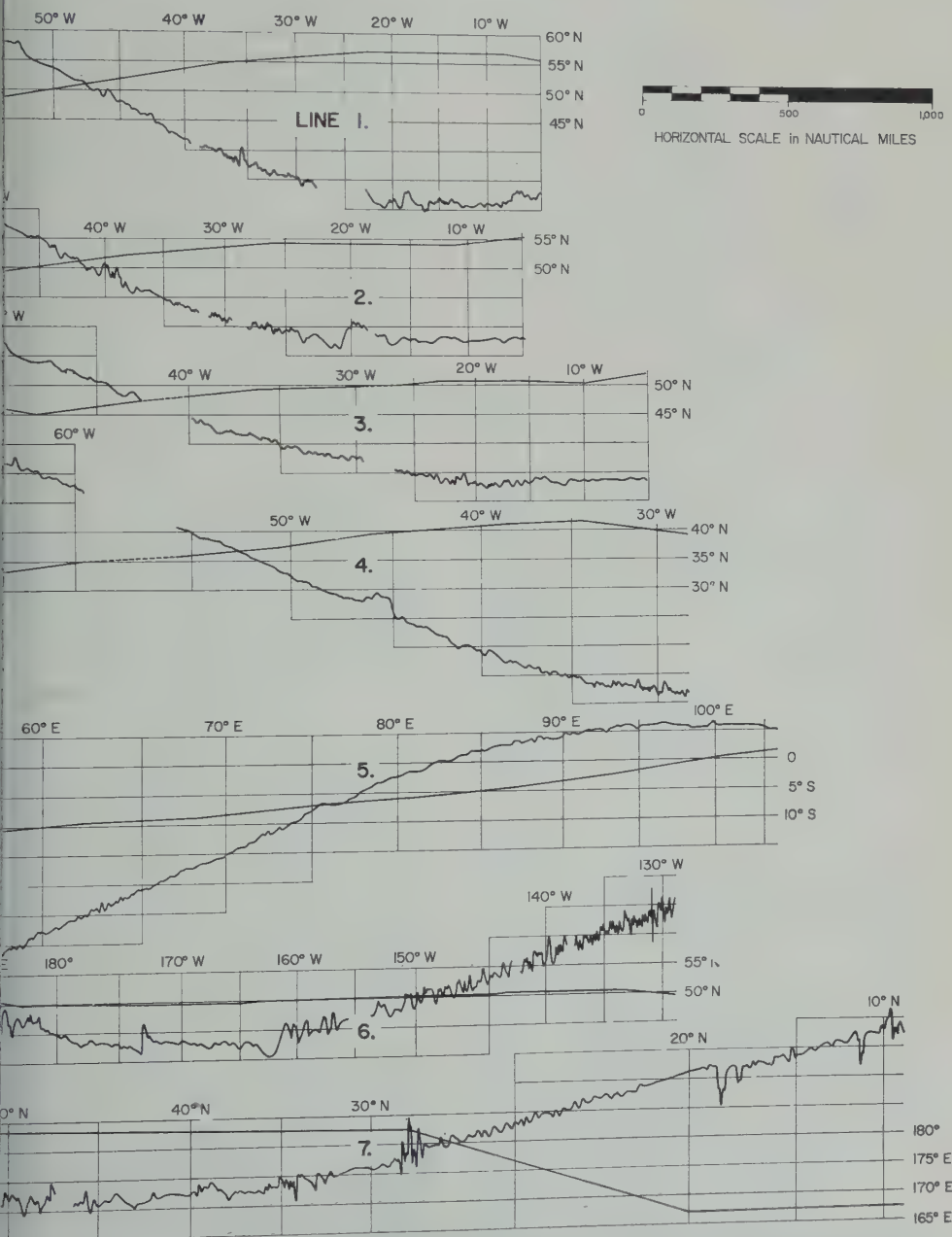


Fig. 2. Profiles showing both position and nondipole field along track lines 1 through 7. The horizontal scale is linear in distance.

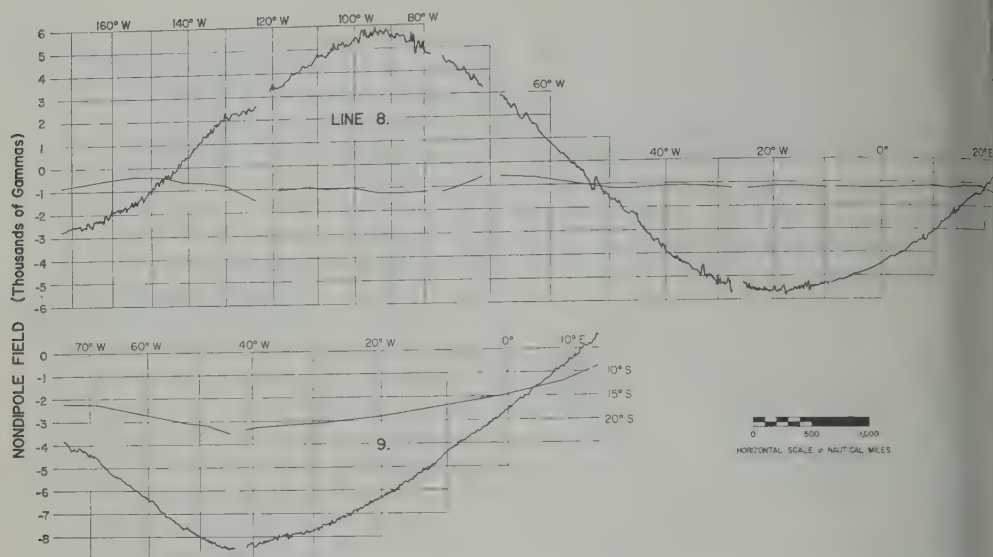


Fig. 3. Profiles showing both position and nondipole field along track lines 8 and 9. The horizontal scale is linear in distance.

2, and 3 were taken from project Magnet, conducted by the U. S. Hydrographic Office at altitudes from 7000 to 19,000 feet; the data line 6 was taken recently by the U. S. Coast and Geodetic Survey with a ship-towed magnetometer; and line 7 was taken in a low-altitude flight made in connection with project Volcano in 1947 [Aldredge and Keller, 1949].

Only total field intensity was used in this analysis, since it can be continuously recorded in a moving vehicle with greater precision than can the magnetic components.

None of the track lines used were true great-circle routes between the end points. Each one had some minor sudden changes in direction which resulted in sudden changes in the slope of the general magnetic trend along the track line. Such sudden changes in the trend induced undesirable effects which had to be eliminated before the analysis could proceed.

To understand the source of these undesirable effects, consider the effect caused when a survey airplane, flying due north, suddenly turns eastward at middle magnetic latitudes. Because of the centered dipole contribution to the surface field, the magnetic record would be increasing up to the eastward turn and would then level off. Any analysis of the results would indicate the presence of unreal higher harmonics to ac-

count for the sudden change in the recorded magnetic trend.

It was found that track lines with turns were not too drastic could still be used. To this, it was found necessary to subtract the computed total field intensity attributed to the centered dipole from the measured total field intensity. The dipole field was taken to be

$$F_d = 0.311(1 + 3 \cos^2 \theta)^{1/2}$$

where θ is the geomagnetic colatitude and F_d is in gauss. The residual is called the nondipole field. The difference between this usage of the term nondipole field and the more common usage [Bullard, Freedman, Gellman, and Nunn, 1950] should cause no confusion.

The maximum north-south gradient of the centered dipole field reaches a maximum of approximately 9 gammas (1 gamma = 10^{-5} gauss) per mile at 54.3° geomagnetic colatitude. The vertical gradient in the total field intensity of the centered dipole field varies from 4.5 gammas per 1000 feet at the geomagnetic equator to twice this value at the poles. The altitude is reasonably constant along any particular track line, but it varied from sea level to 20,000 feet for different track lines. The principal differences noted between high- and low-altitude track lines was the attenuation

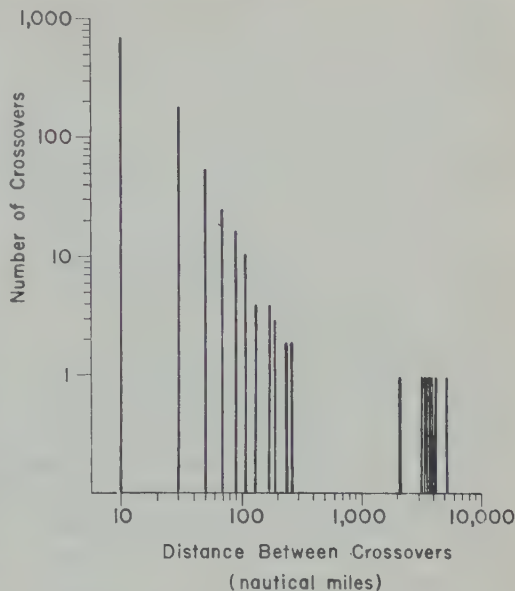
wavelength anomalies at the higher as would be expected.

uble was still observed in the vicinity turns in the track lines even after the centered dipole had been subtracted. occurred in a relatively high gradient associated with one of the major foci of pole field [Bullard, Freedman, Gell-Nixon, 1950], a difficulty exactly like described above for the centered dipole and arise. When such a condition occurred the track line was either rejected or a adjustment was made to bridge the gap. me types of analysis it may prove able to subtract the major features of pole field in addition to the dipole at least the eccentric dipole field from ed data, before the residual anomalies d. This was not done here since to do remove parts of the effect under study. part of the nondipole-field structure account for the major part of the long-h structure.

ack lines and the corresponding nondlds are shown in Figures 1, 2, and 3. k lines shown in Figure 1 are shown in ail in Figures 2 and 3 where they are s straight line segments as a function le and longitude. They are distorted in vay as to preserve the corresponding or the nondipole field as a linear func-distance along the track line.

ures 2 and 3 the heavy solid curves the observed nondipole total-field in-long the indicated track lines. In the a smooth curve, not shown here, was e represent the mean of the observed e field. A stiff spline curve which would ly a minimum of curvature was used. found that the degree of smoothing dded to obtaining a running mean curve each point represents a mean over a length centered on the point in ques-

the most obvious examples of an ir-y in the pattern having been caused by in direction of the track line can be he long composite track line 8 from Midway Island to Athens, Greece, in The track line goes a little north of a le path between Midway Island and cisco in the vicinity of San Francisco.



correspond very closely to the amplitude of a harmonic component because the analysis was done without regard for the amplitude of the curve between crossovers.

The results shown in Figure 4 indicate, for all the track lines considered together, 695 cases having crossover distances between 0 and 20 nautical miles (n.m.), with a successively smaller number of cases as the crossover distance increases until there is no more than one case in any 20-n.m. interval.

The most striking feature of the results shown in Figure 4 is the large gap in the histogram. There is a short crossover distance (short-wavelength) grouping with no cases exceeding 260 n.m. and a long crossover distance (long-wavelength) grouping with no cases shown less than 2100 n.m. (note the logarithmic scale in Fig. 4). There is nearly a factor of 10 between the longest crossover distance in the first grouping and the shortest crossover distance in the second grouping. This large gap in the crossover distance suggests completely different sources for these two groupings. If the difference in sources is related mainly to depth of source, as is discussed later, the above-mentioned gap would imply that there is a depth zone in which magnetic sources are forbidden.

The results shown by the short-wavelength grouping correspond very closely to the results found by *Serson and Hannaford* [1957].

It is realized that this simple analysis does not prove conclusively that there are no small anomalies having wavelengths in the region of the gap referred to above. It can be said quite definitely, however, that the smooth curves do not appear to contain any sizable departures from a more idealized smooth curve and that if anomalies with intermediate wavelengths do exist they are of much smaller amplitude than the bulk of the anomalies that make up the two groupings already discussed.

From the foregoing it is quite clear that in making world magnetic charts the cartographers should strive to include all the long-wavelength anomalies typified by the second grouping in Figure 4, but should ignore the short-wavelength anomalies. This can be done by smoothing continuous data over a distance of 200 n.m. as suggested by *Serson and Hannaford* [1957].

Interpretation of data. The simplest interpretation of the data presented here is that the

long-wavelength anomalies are associated with sources near the core-mantle boundary, short-wavelength anomalies are associated with shallow sources in the part of the crust which is below the Curie temperature, and the middle is a forbidden region for magnetic sources.

This idea has been explicitly stated or implied by many authors [see *Elsasser, 1948; Bullard, Freedman, Gellman, and Nixon, 1950; Henderson and Zietz, 1958*]. The data presented here in the form of very long magnetic profiles are submitted as additional graphic evidence which clearly supports this point of view.

Chatterjee [1956] has suggested that the structure of the earth's magnetism may be in the crust which has a very high permeability caused by high pressure and relatively high temperatures (below Curie point). It is very difficult to understand how such a model could possibly account for the two distinctly different wavelength groups which are found experimentally. If all of the anomalies stem from a common depth zone, more of a continuous distribution of wavelengths would be expected. It would be difficult to explain the westward drift [Elsasser, 1948; Bullard, Freedman, Gellman, and Nixon, 1950] with this model.

There can be little argument against the assumption that the short-wavelength group anomalies (0 to 260 n.m. crossover distances) caused by sources within the earth's crust. The majority of these short wavelengths can be explained by point sources above the depth at which the Curie temperature occurs. The broad anomalies of this group can be explained by broad, deeply buried lithologic structures involving varying depths or changes in permeability.

An estimate of the depth to the sources of the long-wavelength group (2100 to 5200 n.m. crossover distances) has been obtained through the use of simple theoretical models. Three magnetic dipoles were used to approximate the observed field along line 8 (see Fig. 3). Since the magnitude of the total field of a centered dipole was subtracted from the magnitude of the observed field in the experimental data reported in this paper, a similar procedure was tried in the model computations. First, the effect of the observed dipole was added vectorially to the field of a centered dipole. Then the magnitude of the centered dipole field was subtracted from the magnitude of this vector sum. It was found

that the difference between the vector algebraic additions, at the latitude of line 9, was very small. This permitted a similar in the model computations in which that of the centered dipole could be ignored. Negative dipoles were placed under the minimums and a positive dipole under the maximum. The magnitudes of the various dipoles were made proportional to the associated maximum and minimum values of the observed field giving this control on the magnitudes and positions of the dipoles leaves only their depth to be determined. Curve matching techniques were used to estimate of the depth.

Figure 5 shows a series of computed curves compared with a smoothed portion of line 8. The depth parameter p is the radial distance from the earth's center to the dipole normalized with respect to the earth's radius. It is apparent that the computed curves are fairly sensitive to the value of p and that the best fit occurs when p is about 0.5. This indicates that the anomalies observed with the very long wavelengths can be explained by sources near the core-mantle boundary. This agrees with the dynamo theory, which requires a number of current loops near the core-mantle boundary [Inglis, 1955].

The current loops of the dynamo theory are approximated by the dipoles used here. Van Laue [1940] closely approximated the major portion of the nondipole field by using 14 radial dipoles at the core-mantle boundary. These demonstrations, of course, do not prove conclusively that the long-wavelength variations can be accounted for by sources nearer the core-mantle boundary, but such explanations seem very forced and unlikely.

The preceding discussion indicates that the magnetic field of the earth, as measured, can be approximated very easily by a set of sources in the crust and a second set near the core-mantle interface.

Project Magnet. Project Magnet, being conducted by the U. S. Hydrographic Office, is a part of the United States contribution to the world magnetic survey. Within the past seven around-the-world flights have been completed as a part of this project. These data present a unique opportunity to study the nature of the magnetic anomalies.

Unfortunately that the project Magnet cannot be made along tracks of equal

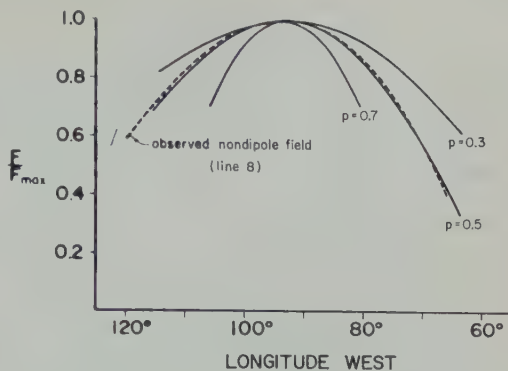


Fig. 5. Comparison of the observed nondipole field along line 8 with the field of three radial dipoles. The positions and magnitudes of the dipoles were determined from the positions and magnitudes of the maximums and minimums along line 8. p is the distance from the earth's center to the dipoles normalized to the earth's radius.

latitude, for then the results could be used directly in a harmonic analysis without modification. In spite of the somewhat circuitous routes which are required by aircraft limitations, the around-the-world flights should prove very helpful in a further investigation of forbidden wavelengths as discussed in this paper and in a discussion of harmonic analysis of the earth's field.

The USC&GS and the Hydrographic Office are cooperating on an analysis of the straightest available composite around-the-world flights. The centered dipole field is being subtracted first, as described in this paper, to remove unreal wavelengths which may appear because of north-south wander of the track line through the main dipole field. Digital values for the total field intensity will be picked off the residual continuous around-the-world record every 3 miles. A Fourier analysis of these values will then be made on a high-speed computer. Harmonic coefficients up to the 2000 order will be obtained. This range of analysis should indicate the behavior of the coefficients from the very long wavelengths, corresponding to the broad regional anomalies, up to the very short wavelengths of approximately 10 miles in length, corresponding to very shallow anomalies in the crust of the earth.

REFERENCES

- Allredge, L. R., and Fred Keller, Jr., Preliminary report on magnetic anomalies between Adak,

- Alaska, and Kwajalein, Marshall Islands, *Trans. Am. Geophys. Union*, 30, 494-500, 1949.
- Bullard, E. C., Cynthia Freedman, H. Gellman, and Jo. Nixon, The westward drift of the earth's magnetic field, *Phil. Trans. Royal Soc. London, A*, 243, 67-92, 1950.
- Chatterjee, J. S., The crust as the possible seat of earth's magnetism, *J. Atmospheric and Terrest. Phys.*, 8, 233-239, 1956.
- Elsasser, W. M., The earth's interior and geomagnetism, *Revs. Modern Phys.*, 22, 1-35, 1950.
- Henderson, J. R., Jr., and I. Zietz, Interpretation of an aeromagnetic survey of Indiana, *U. S. Geol. Survey Profess. Paper 316-B*, 1958.
- Inglis, D. R., Theories of the earth's magnetism, *Revs. Modern Phys.*, 27, 212-248, 1955.
- Knapp, D. G., The new isogonic chart of the United States, *Terrest. Magnetism and Atmospheric Elec.*, 47, 53-62, 1942.
- McNish, A. G., Physical representations of the geomagnetic field, *Trans. Am. Geophys. Union*, part 2, 287-291, July 1940.
- Serson, P. H., and W. L. W. Hannaford, A statistical analysis of magnetic profiles, *J. Geophys. Research*, 62, 1-18, 1957.
- Vestine, E. H., L. Laporte, I. Lange, and W. Scott, The geomagnetic field, its description and analysis, *Carnegie Inst. Wash. Publ.* 5, 1947.

(Manuscript received July 20, 1961.)

Effects of Moderate Stresses on Directions of Thermoremanent Magnetization

JOHN W. KERN¹

*University of California
Berkeley, California*

Abstract. Experiments designed to test the effects of directed stress on the thermoremanent magnetization (TRM) of igneous and metamorphic rocks are described. Basalt and andesite specimens exhibited no anomalous directions of TRM when cooled from 600°C under uniaxial stresses of up to 350 bars. Metamorphic rocks with preferred crystallographic orientations were found to exhibit anomalous directions of induced TRM after such treatment. These anomalous TRM directions were apparently related to the lineation of the specimens rather than applied stress. Residual magnetizations of the metamorphic specimens were found to approach the direction of the magnetizing field upon progressive alternating-field demagnetization.

Introduction. In a series of isothermal experiments on metamorphic gneisses from the Adirondack Belt in the Adirondack Mountains, Graham, Buddington, and Balsley [1957] observed marked changes in intensity and direction of magnetization by changing the axial stress on specimens from 24 to 182 bars. They concluded that magnetostriction effects are of major importance in interpreting measurements of remanent magnetization, noting that although many rocks acquire magnetization under directed stress, magnetic measurements are made under relatively stress-free conditions. Thus the directions of magnetization as measured may not represent the original *in situ* magnetizations of the specimens, and the measured directions may not reflect the direction of the magnetic field at the time of the original magnetization.

In paleomagnetic research, the primary aim is to determine the information embodied in the magnetization of a rock specimen, primarily the direction of magnetization associated with its formation. The idea that complicated rocks composed of magnetic material with homogeneous magnetization or under directed stress can be immediately rejected. From the discussion of stability of remanent magnetization and the use of alternating-field and thermal-demagnetization procedures in the following papers, we can see the information embodied in different com-

ponents of magnetization may differ in reliability. Further, establishing criteria for stability of remanent magnetization on the basis of alternating-field or thermal-demagnetization procedures leads to the formulation of reliability criteria on the same basis.

As will be shown in following papers, criteria based on these procedures can be established for evaluating the effect of stress on remanent and induced magnetization even for heterogeneous systems.

In view of the higher stability of thermoremanent magnetization (TRM) in alternating-field (a-f) demagnetization procedures, and the comparatively low stability of the remanent magnetizations of the metamorphic rocks used by Graham, Buddington, and Balsley, drawing conclusions from isothermal experiments on these rocks applying to TRM in general is a somewhat questionable procedure. The most recent structural theory of TRM is that of Verhoogen [1959], who takes into account information gained in a-f demagnetization experiments and in the negative results of the experiments conducted to check the hypothesis, described in the next paper, that stress does affect TRM.

Stott and Stacey [1959, 1960] reported on a series of experiments designed to test the hypothesis that directed stress can have effects on the direction of TRM in rocks of interest in paleomagnetism. Uniaxial stress was applied to rock specimens cooling from high temperatures in the earth's magnetic field. Directions of mag-

¹ Present address: Rand Corporation, Santa Monica, California.

netization of these specimens were compared with those of specimens cooled in the earth's field with no applied stress. Stott and Stacey concluded that magnetostriction effects on the direction of TRM of igneous rocks are negligible and that effects significant in paleomagnetism are improbable.

Recent experiments which may indicate irreversible effects of stress on directions of thermoremanent magnetization are those of Hall and Neale [1960]. Departures of the order of 5° from TRM directions in an unstressed state were obtained in a series of experiments in which a variety of rocks was used. The amount of departure for a given specimen was roughly proportional to the applied stress and inversely proportional to grain size. The specimens used were magnetically anisotropic under thermal magnetization; that is, the directions of TRM did not coincide with the direction of the magnetizing field. Stress-dependent rotations of TRM for this type of specimen may be correlated with plastic deformation, since the treatment involves heating to 600°C and applying pressures of over 100 bars. If plastic deformation of the nonmagnetic matrix occurs, size-dependent rotations of elongated magnetic mineral grains might be expected. In order for stress effects on TRM to be demonstrated, specimens must be magnetically isotropic with equidimensional magnetic mineral grains. This would ap-

pear more likely to be the case if rapidly cooled basalt flows are sampled, rather than injected or metamorphic rocks. Practically, the problems of deformation and magnetostriction are inseparable for anisotropic specimens.

Graham, Buddington, and Balseley [1958] pointed out that the results of Stott and Stacey do not exclude effects of a time-dependent structure, such as those which might be exhibited by a magnetic crystal phase exsolving from a ferrimagnetic solid solution over a long period of time. Unfortunately, time spans of the duration encountered in nature cannot be duplicated in the laboratory, and hence it seems unlikely that experiments will resolve such questions.

Stress effects on TRM of igneous rocks. A series of experiments conducted by the author produced additional negative evidence for magnetostriction effects on the direction and intensity of TRM. Saturation-magnetization curves were determined for each rock used, giving the Curie temperature for the magnetic minerals present and allowing the composition of the magnetic constituents to be estimated (see Table 1). Specimens were placed in the apparatus, oriented by means of reference marks, heated to temperatures above their respective Curie points, and cooled under selected uniaxial stresses.

Successive specimens of the same rock were subjected to axial stresses of different magni-

TABLE 1. Igneous Rock Composition and Magnetization Data

Symbol	Rock Description	Curie Point, °C	Sample Saturation Magnetization,* emu/g	Magnetic Material,† Wt %	Inferred Mineral Composition	Number of Samples	Maximum Applied Stress, bars	Mean Deviation of Zero Stress Sample Magnetization, degrees
S1RD	Olivine basalt	515	48	5.2	MI‡	5	175	1.2
S1RE	Olivine basalt	535	50	1.4	MI	6	220	0.9
S1RF	Andesite	540	51	1.9	MI	3	220	1.0
S1RG	Olivine basalt	495	46	2.4	MI	5	220	1.4
S1RI	Quartz basalt	520	48	0.8	MI	5	260	1.4
S1RJ	Andesite	525	49	2.5	MI	5	350	1.1
S1RK	Olivine basalt	515	17	2.0	MU§	25	350	1.7

* From data and methods of Vincent, Wright, Chevallier, and Mathieu [1957], computed from Curie point and characteristic form of saturation magnetization curve.

† Computed from force exerted on saturated sample at room temperature in a magnetic field of known gradient.

‡ MI—Magnetite-Ilmenite solid solution series.

§ MU—Magnetite-Ulvöspinel solid solution series.

2. Descriptions and Iron and Titanium Oxide Compositions of Adirondack Metamorphic Rocks Used in TRM-Stress Experiment

Description	Iron and Titanium Oxides (wt % of Rock)				
	Magnetite Fe ₃ O ₄	Hematite Fe ₂ O ₃	Ilmenite FeO·TiO ₂	Rutile TiO ₂	Total
Amphibolitic metagabbro gneiss	0	0	0	0	0
Phacoidal hornblende granite gneiss	1.09	—	0.55*	—	1.64
Pyroxene diorite	0	0.42	3.75	0.11	4.28
Hornblende-andesine amphibolite	0.11	1.66	1.17	0.38	3.32
Contaminated hornblende microcline granitic gneiss	5.09	0.56	0.77	0.17	6.59
Partly chloritized biotite-hornblende-andesine gneiss	0.09	0.30	0.05	0.01	0.45
Hornblende microperthite granite	0	0.05	0.43	0.01	0.49
Hornblende microperthite granite	0	0.05	0.43	0.01	0.49
Amphibolite gneiss	0.03	0.05	0.42	0	0.50
Microperthite alaskite	1.02	0.07	0.07	0	1.16
Hornblende microcline-plagioclase granite gneiss	1.97	—	0.11*	—	2.08
Microcline granite gneiss, accessory pyroxene, hornblende, sphene	4.87	—	0.70*	—	5.57

* This number applies to the total of Fe₂O₃, FeO·TiO₂ and TiO₂.

The specimen magnetizations were measured by means of a rock generator magnetometer of the type described by Doell [1955] and [1959]. A comparison of directions of TRM of these specimens gives a direct indication of the effect of axial stress.

In the exception of the experiment denoted by 1, the applications of axial stress and magnetic field were continuous during cooling to low temperature. In S1RK, an olivine basalt, a very nearly linear cooling saturation-magnetization curve was employed in an attempt to establish possible stress effects on low-temperature TRM. The linear character of the saturation-magnetization curve for this rock permitted application of stress and field over a wide range of temperature intervals in cooling with the result of equal resulting partial thermoremanent magnetizations. This resulted in less stringent requirements on the temperature control in the experiment. The experimental error was nevertheless increased owing to variation in the temperature at the end points of the interval over which stress and field were applied.

The over-all experimental error can be estimated as about $\pm 2^\circ$. The results presented are variations δ in angle between the direction of magnetization and the stress axis as com-

puted from the zero stress specimen magnetization in each set. These results are shown in Table 1.

As can be seen in Table 1, the mean deviations of δ do not exceed the estimate of the over-all experimental error. Further, no trends in departure from the zero stress direction are evident. The applied stress approached the breaking strength of the rocks at the temperature reached; in some cases the specimens were fractured. Thus thermoremanent magnetization of the rocks employed in this series of experiments appears stable under stresses of geologic importance. Most of the rocks contained minerals of the magnetite-ilmenite series as determined by the methods of Vincent, Wright, Chevallier, and Mathieu [1957] from the saturation-magnetization data (see Table 1). In the experiments of Graham, Buddington, and Balsley [1957], members of this series exhibited fairly large isothermal stress effects.

Stress effects on TRM of Adirondack metamorphic rocks. Samples of Adirondack metamorphic rocks, supplied by J. R. Balsley of the Geophysics Branch of the U. S. Geological Survey, were subjected to treatment of the type described in the preceding section. These are rocks from the same belt of gneisses as those used by Graham, Buddington, and Balsley

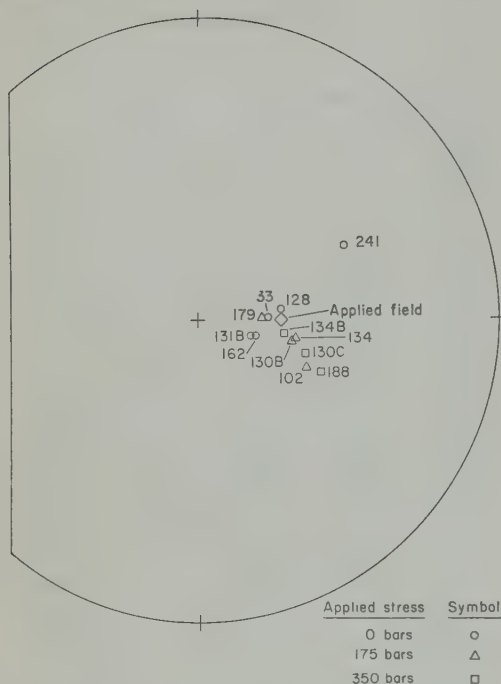


Fig. 1. TRM directions of Adirondack specimens, displayed on Schmidt net. The stress axis is taken as the axis of the projection and the center of the net.

[1957] in their experiments on isothermal magnetostriction effects. Descriptions and iron-titanium oxide compositions are given in Table 2.

Three groups of specimens were used: the first cooled in an applied field with no stress, the second with applied stress of 175 bars, the third with 350 bars. The results are presented in the form of projection of the directions of magnetization on a Schmidt net. The directions of magnetization appeared unstable in the earth's field, as shown by measurements made 48 hours later. The results of these measurements are shown in Figure 1. Owing to fracture, retest under different applied stress was not possible for most of the specimens. Specimen 241 departed most markedly from the direction of the applied field, no stress being applied in the first treatment. The direction of magnetization for specimen 241 was very nearly the same following retest with 175 bars applied stress. A third treatment, in which the specimen was rotated 90° with respect to the applied field about the stress axis and no stress was applied,

gave a different direction of magnetization with respect to the applied field. Figure 2 shows the directions of magnetization relative to the specimen and the stress axis. Since applied stress has no effect and the relation of the direction of magnetization to that of the applied field depended on the orientation of the field with respect to the specimen, the conclusion may be drawn that the direction of TRM is affected by anisotropic distribution of crystallographic axes in the magnetic mineral constituents of the rock.

The Adirondack metamorphics exhibit lineation (preferred crystallographic orientations), as found by *Balsley and Buddington* [1955] and reported by *Graham, Buddington, and Balsley* [1957]. The directions of magnetization of the magnetite-bearing specimens from this region have a mean very close to the mean lineation. *Daly* [1959] and *Grabovsky and Brodsky* [1958] have shown that metamorphic rocks do not necessarily acquire TRM parallel to the direction of an applied field. Such rocks exhibit anisotropic susceptibility. This property may prove useful in testing for possible preferred crystal orientations which may influence the

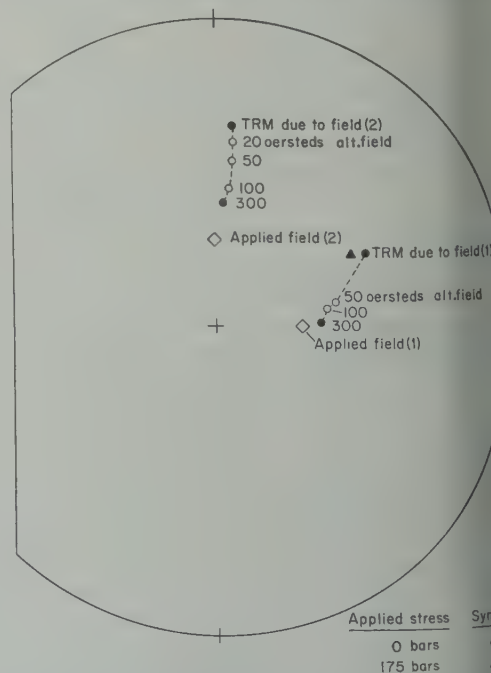
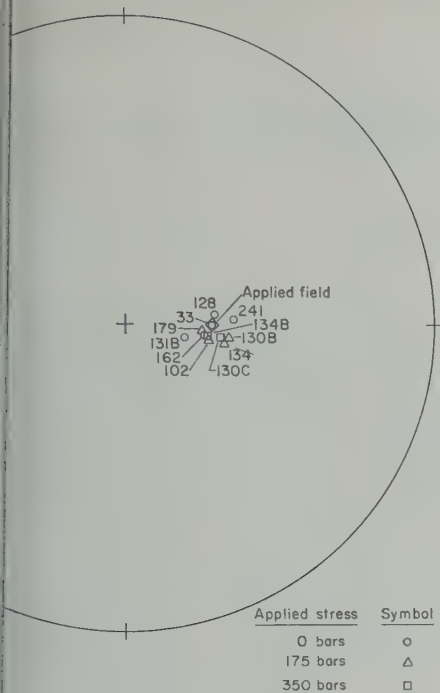


Fig. 2. TRM directions of Adirondack specimen 241 under different applied fields and stress with results of a-f demagnetization.



TRM directions of Adirondack specimens following 300-oersted a-f demagnetization.

of paleomagnetic measurements. Alternate-field demagnetization procedures were used on the treated Adirondack metamorphic rocks. The results are shown in Figure 3. In all cases, the agreement between the direction of magnetization and that of the applied field improved with successive applications of increasing alternating fields. This can be interpreted as resulting from a demagnetization process associated with crystallographically oriented grains. Thus the demagnetized rocks retain magnetizations associated with grains exhibiting no preferred directions of magnetization. This supports the model inferred by Verhoogen [1959] for stable high-temperature TRM. In this model TRM is ascribed to the effect of crystal structure dislocation with preferred orientation in space. The results also support the validity of the argument that stable TRM is directly related to reliability of paleo-

magnetic data for a given component of magnetization.

Acknowledgments. This work was done at the University of California, Berkeley, California, aided by grants from the American Petroleum Institute and the Institute of Geophysics of the University of California, and was partially carried out under a Fellowship of the American Chemical Society Petroleum Research Fund.

REFERENCES

- Balsley, J. R., and A. F. Buddington, Remanent magnetism of the Russell belt of gneisses, *Advances in Physics*, 6(23), 317-322, 1955.
- Cox, A. V., The remanent magnetization of some Cenozoic rocks, Doctoral dissertation, University of California, Berkeley, 193 pp., 1959.
- Daly, Lucien, Sur l'anisotropie magnétique dans les roches déformées, et la nature de leur aimantation rémanente naturelle, *Compt. rend.*, 248, 2614-2616, 1959.
- Doell, R. R., Remanent magnetism in sediments, Doctoral dissertation, University of California, Berkeley, 156 pp., 1955.
- Grabovsky, M. A., and S. Yu. Brodskaya, Normal magnetization and thermomagnetization of anisotropic rocks, *Izvest. Akad. Nauk SSSR Seria Geofiz.*, 8, 977-988, 1958 (Engl. Edit. Amer. Geophys. Union, Nov. 1959).
- Graham, J. W., A. F. Buddington, and J. R. Balsley, Stress-induced magnetizations of some rocks with analyzed magnetic minerals, *J. Geophys. Research*, 62(3), 465-474, 1957.
- Graham, J. W., A. F. Buddington, and J. R. Balsley, Magnetostriction and paleomagnetism of igneous rocks, *Nature*, 183, 1318, 1959.
- Hall, J. M., and R. N. Neale, Stress effects on thermoremanent magnetization, *Nature*, 188 (4753), 805-843, 1960.
- Stott, P. M., and F. D. Stacey, Magnetostriction and paleomagnetism of igneous rocks, *Nature*, 183, 384-385, 1959.
- Stott, P. M., and F. D. Stacey, Magnetostriction and paleomagnetism of igneous rocks, *J. Geophys. Research*, 65, 2419-2424, 1960.
- Verhoogen, J., The origin of thermoremanent magnetization, *J. Geophys. Research*, 64(12), 2441-2449, 1959.
- Vincent, E. A., J. B. Wright, R. Chevallier, and S. Mathieu, Heating experiments on some natural titaniferous magnetites, *Mineral. Mag.*, 31, 624-655, 1957.

(Manuscript received May 5, 1961; revised August 18, 1961.)

The Effect of Stress on the Susceptibility and Magnetization of a Partially Magnetized Multidomain System

JOHN W. KERN¹

*University of California
Berkeley, California*

Abstract. An analysis of the effects of directed stress on magnetization and susceptibility is presented, with special attention to the problems of rock magnetism. A model is proposed for partial magnetization of an isotropic multidomain magnetic system. The behavior of the model system in response to applied stress is analyzed. It is found that the response of the system to applied stress, given by the ratios of magnetization and susceptibility to their respective initial values, is simply related to the saturation magnetization and magnetostriction parameters for the system and to the initial susceptibility. The behavior of these two ratios is approximately the same. This behavior can be approximated very closely by the function $a/(S + a)$, where a is defined as the 'virtual initial stress' and S is the normal deviatoric component of the applied stress. The value of a is given by $J^2/3\lambda\chi$, where J is the saturation magnetization, λ the saturation magnetostriction, and χ the initial susceptibility. The theoretical results are compared with isothermal experiments by Grabovsky and Parkhomenko for magnetite and titanomagnetite, and the agreement is found to be quite good. The virtual initial stress parameter can be calculated for high temperatures using Kneller's experimental data for nickel. The conclusion is drawn that the stress response is regular with temperature, and it has no strong effects even where the magnetization vanishes at the Curie point.

Introduction. The effects of stress on rock magnetization are of crucial concern for both magnetism and paleomagnetism [Graham, Barton, and Balsley, 1957, 1959; Stott and Barton, 1959, 1960]. This problem can be dealt with by the development of stress stability criteria for magnetization. The analysis given here deals with the effects of applied stress to measure properties of physical magnetic systems, such as the saturation magnetization and magnetostriction parameters and the initial susceptibility of a magnetic material. It applies to partial or total magnetization.

The development given here is an outgrowth of an analysis of the experiments of Grabovsky and Parkhomenko [1953] and Graham, Budington, and Balsley, [1957], which dealt with the effects of uniaxial stress on rock magnetization. A theory developed by Stacey [1958], based on the experimental results of Graham, Barton, and Balsley [1957], predicts the error of one of Graham's magnetite specimens gives erroneous results for two others. The model proposed by Stacey appears to be either incorrect or too simple to account for the results at the Rand Corporation, Santa Monica, California.

the behavior of complicated magnetic systems, such as the metamorphic rocks used by Graham and others.

The assumptions necessary for the model introduced here are stated at the outset. Many of the assumptions introduced are required for expression of the results in closed form. Tensor notation is used in order to obtain general constitutive relations as a starting point. The question of reversibility of stress effects is not dealt with here.

General. A partially magnetized multidomain system is assumed herein to have the following properties:

(1) The system consists of a crystal aggregate with randomly oriented crystal axes; hence it will exhibit no crystalline anisotropy.

(2) The grains making up the aggregate are multidomain grains.

(3) Domain walls are subject to displacement under the action of suitable driving forces.

(4) The magnetization of the system is given by the vector sum of the magnetizations of the constituent domains.

(5) The domain magnetizations may be grouped into sets characterized by homogeneous directions of magnetization.

(6) Each of these sets exhibits the property of susceptibility; that is, components of magnetization change in response to a varying external applied field.

(7) The system is homogeneous and isotropic with respect to the following properties: (a) saturation magnetostriction parameter λ ; (b) saturation magnetization, J ; and (c) initial susceptibility, χ_0 , of the system at any state.

The term *multidomain* used in (2) above means that each grain consists of a number of regions, each spontaneously magnetized to saturation in some direction and separated by so-called domain boundaries or walls. The domain structure depends on the magnetic anisotropy of the grain, which determines the possible directions of spontaneous magnetization. Magnetic properties of such a grain can be analyzed in terms of changes in this hypothetical internal structure.

This model is based on the domain theory of Weiss [1907]. The moving domain boundary was proposed by Bloch [1932] and Becker [1932, 1933]. Becker showed the moving boundary to be especially important for processes occurring in low and medium fields.

The present treatment applies to heterogeneous systems as well as to homogeneous systems if the following assumptions are made.

(1) Any given heterogeneous system can be subdivided into a number of homogeneous subsystems.

(2) Magnetic interactions between constituent subsystems can be neglected.

The variables used to construct a free energy function for a magnetoelastic system, expressed in cartesian tensor notation [Jeffreys, 1931] are: (1) the magnetoelastic strain tensor, ϵ_{ij} ; (2) the stress tensor, σ_{ij} ; (3) the magnetization vector, j_i ; and (4) the ambient magnetic field vector, H_i . These can be combined in an energy function with the differential form

$$dg = \epsilon_{ij} d\sigma_{ij} + j_i dH_i \quad (1)$$

This function is the change in potential energy for a magnetoelastic system in which the strain and magnetization are functions of the magnetic field. Bozorth [1951] and Stacey [1958] have applied this relation in a more special form, considering only normal stresses. Here $\epsilon_{ij} d\sigma_{ij}$ represents the increment in strain energy for an infinitesimal stress $d\sigma_{ij}$, due to a magnetic strain

ϵ_{ij} , while $j_i dH_i$ is the increment in magnetic potential energy due to an infinitesimal change in the magnetic field. Clapeyron's theorem, relating strain energy to the equilibrium stress and total strain of an elastic medium, ensures that strains arising from magnetization contribute the same way to strain energy as those arising from stresses [cf., Sokolnikoff, 1956, p. 8]. Equation 1 is applicable only to systems which are homogeneous in all the above variables. Therefore, it will be applied here to sets of domains characterized by homogeneous directions of magnetization as mentioned above. Note that the Einstein convention of summation over repeated indices is used throughout. From this we can write an important Maxwell's relation

$$(\partial \epsilon_{ij} / \partial H_k)_\sigma = -(\partial j_k / \partial \sigma_{ij})_H$$

This relation is of central importance in the following development, implying reversibility of magnetization and magnetostriction.

Susceptibility of a homogeneous domain. A population of single domains, associated with multidomain grains and characterized by the same direction of magnetization, will exhibit susceptibility as defined by

$$\chi_{mn} = (\partial j_m / \partial H_n)$$

This is due to the transition of domain magnetization by translation of domain walls from one (or preferred) directions of spontaneous magnetization into other directions more favorably located with respect to the applied field. Thus, the system approaches a state of lower internal energy by a change in direction of magnetization with respect to the applied field. This mechanism gives rise to a continuous change in magnetization as the applied field is changed. Here only reversible changes of magnetization will be considered.

For isotropic magnetostriction, the stress tensor takes a form such that the extension is in a direction given by direction cosines l_i

$$\epsilon = \epsilon_{ij} l_i l_j = 3\lambda/2[(\alpha_i l_i)^2 - 1/3]$$

where α_i are the direction cosines of the magnetization vector j_i , and λ , the saturation magnetostriction coefficient, is defined by extension $\delta l/l$ in the direction of saturation magnetization in very high fields. Now the stress tensor may be written as $\sigma_{ij} = S_{ij} + \frac{1}{3}\epsilon \delta_{ij}$ where S_{ij} is the deviatoric stress tensor and

an octahedral or hydrostatic stress. In this relation, we can write the magneto-coupling term in (1), noting that the term in the magnetoelastic coupling can be written,

$$\epsilon_{ij} d\sigma_{ij} = e_{ij} dS_{ij} \quad (5)$$

where e_{ij} is the deviatoric component of ϵ_{ij} , and S_{ij} are the deviatoric stress components alone. The variation of the energy function, in complete expression is

$$dg = e_{ij} dS_{ij} - j_k dH_k \quad (6)$$

where the normal component of deviatoric stress in the direction given by the direction i is given by $S_i = S_{ij}l_jl_i$.

From (6) we can write another Maxwell's

$$\left(\frac{\partial e_{ij}}{\partial H_k}\right)_{S_i} = -\left(\frac{\partial j_k}{\partial S_{ij}}\right)_H \quad (7)$$

For a homogeneously magnetized set of domains, the direction cosines are given by

$$\alpha_p = j_p/J \quad (8)$$

where J is the saturation magnetization per unit volume for the material, and j_p is the p -axis component of the magnetization vector ($p = 1, 2, 3$). Therefore, from (4) and (8), it follows that

$$\left(\frac{\partial j_k}{\partial S}\right)_H = -\frac{3\lambda}{J^2} (j_p l_p) \left(\frac{\partial j_n}{\partial H_k}\right)_S l_n \quad (9)$$

and the derivative of (9) with respect to applied field, it can be shown that

$$\left(\frac{\partial \chi}{\partial S}\right)_H = -\frac{3\lambda}{J^2} \left[\chi^2 + j \left(\frac{\partial \chi}{\partial H}\right)_S \right] \quad (10)$$

where H and S are understood to be in the same direction, χ is the susceptibility, and j is the magnetization in the direction of the applied S . To obtain a differential equation for the variation of susceptibility with applied S , we make use of continuity requirements for j and the derivatives of j with respect to H and S . The existence of relations of the type

$$\frac{\partial}{\partial H} \left(\frac{\partial \chi}{\partial S}\right)_{H,S_i} = \left[\frac{\partial}{\partial S} \left(\frac{\partial \chi}{\partial H}\right)_{S,H} \right]$$

and from (10) we can derive

$$\left(\frac{\partial^2 \chi}{\partial S^2}\right)_H + 6 \left(\frac{3\lambda}{J^2}\right) \chi \left(\frac{\partial \chi}{\partial S}\right)_H - 4 \left(\frac{3\lambda}{J^2}\right)^2 \chi^3 = 0 \quad (11)$$

This is the desired equation relating susceptibility to stress at constant field.

The general solution of this second-order differential equation is

$$\chi = \frac{C_1 \chi_0}{sS + 1} + \frac{C_2 \chi_0}{2sS + 1} \quad (12)$$

where use has been made of the zero stress conditions: at $S = 0$, $\chi = \chi_0$, $j = j_0$, $(\partial \chi / \partial H)_S = f_0$, and where $s = 3\lambda \chi_0 / J^2$. C_1 and C_2 remain to be determined from the initial conditions.

Expressed as the ratio of final to initial susceptibility, r_χ , equation 12 gives the variation of susceptibility with stress as

$$r_\chi = \frac{\chi}{\chi_0} = \frac{C_1}{sS + 1} + \frac{C_2}{2sS + 1} \quad (13)$$

The susceptibility given by (12) can now be substituted into (9) to give the differential equation for the variation of magnetization with applied stress at constant field. For the component of magnetization parallel to the applied stress, we can write this equation as

$$\left(\frac{\partial j}{\partial S}\right)_H = -\frac{3\lambda \chi_0}{J^2} j \left[\frac{C_1}{sS + 1} + \frac{C_2}{2sS + 1} \right] \quad (14)$$

This can be integrated directly to give the magnetization as a function of s , S , C_1 , and C_2 . The result can be expressed as the ratio of final to initial magnetization, r_j . Thus

$$r_j = \frac{j}{j_0} = \left(\frac{1}{sS + 1} \right)^{C_1} \left(\frac{1}{2sS + 1} \right)^{C_2} \quad (15)$$

where j_0 is the initial magnetization in the direction of the applied stress S , j is the magnetization at the stress S , s is as defined above, and C_1 and C_2 are to be determined.

C_1 and C_2 are determined from the initial (zero stress) conditions given above at constant H . From (10) and (12),

$$C_1 = 1 - j_0 f_0 / \chi_0^2 \quad (16)$$

$$C_2 = j_0 f_0 / \chi_0^2$$

It is possible to obtain a simpler representation of C_2 . For induced magnetization from an initial demagnetized state, ferrite magnetization in a weak field can be adequately represented by

$$j = \chi_a H + f H^2 / 2$$

where χ_a and f are constants. This is called Rayleigh's first relation [Neel, 1955]. The associated susceptibility is the first derivative of j with respect to H ; i.e., $\chi = \chi_a + fH$, where $f = (\partial\chi/\partial H)_s$ and χ_a is the initial or zero field susceptibility.

Assuming the above Rayleigh's relation for the system under consideration here, we can show that $jf/\chi^2 = (1 - \chi_a^2/\chi^2)/2$. If we denote the zero stress values of the magnetization, the susceptibility, and the derivative of susceptibility with respect to applied field by j_0 , χ_0 , and f_0 , respectively, C_2 assumes the value $j_0 f_0 / \chi_0^2$. It follows that

$$C_2 = (1 - \chi_a^2/\chi_0^2)/2 \quad (17)$$

where χ_a is the value of susceptibility at zero stress and zero field.

It is apparent that for weak fields, such as the earth's, the constant C_2 has a maximum value of $1/2$; and that for $\chi_0 \cong \chi_a$, $C_2 \cong 0$. Note that for isotropic and homogeneous susceptibility, C_2 is also isotropic and homogeneous. Now for $C_2 \ll 1$, r_x and r_j are very nearly equal. That is

$$r_x = r_j = r = 1/(sS + 1) \quad (18)$$

where s is as given above. This quantity will be used to give a first approximation to the stress dependence of χ/χ_0 and j/j_0 .

The maximum departures of r_x and r_j from a theoretical curve for r are shown in Figure 1.

The quantity s , given by

$$s = 3\lambda\chi_0/J^2 \quad (19)$$

where λ is the saturation magnetostrictive parameter for the material, χ_0 is the initial susceptibility (zero stress at constant field) and J_s is the saturation magnetization for the material, will be called the *stress sensitivity* of the material. In the following discussion, use of the term will in the main refer to the approximate solution given by (18). Note, however, that the stress sensitivity s plays a dominant role in determining the stress-dependent behavior of a magnetic system, particularly in weak magnetic fields.

Curve B is that calculated from (18), and curves A and C are calculated from (13) and (15) with C_2 equal to its maximum value of $1/2$. Comparison of these curves indicates the value of (18) as an approximation to the complete solution given by (13) and (15).

Normal deviatoric stress effects on a partially magnetized multidomain system. The preceding section dealt with the effect of normal deviatoric stress on a multidomain magnetic system homogeneous in saturation magnetization and saturation magnetostriction, and homogeneous and isotropic in susceptibility. The results can be applied directly to a general, partially magnetized system by summation over constituent homogeneous subsystems. Stress-dependent behavior of susceptibility and magnetization

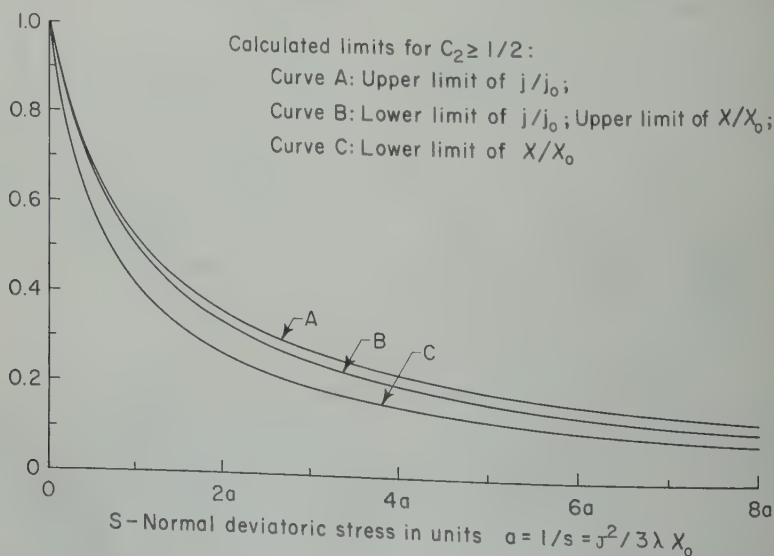


Fig. 1. Theoretical limits for χ/χ_0 and j/j_0 as a function of normal deviatoric stress.

erogeneous magnetic system is obtained on of susceptibilities and magnetizations constituent subsystem weighted according to volume fraction. Thus, for a system having homogeneous subsystems

$$\chi = \sum_i^N v^{(i)} \chi_0^{(i)} r_x^{(i)} \quad (20)$$

$$j = \sum_i^N v^{(i)} j_0^{(i)} r_j^{(i)} \quad (21)$$

is the index of summation, $v^{(i)}$ is the fraction of the i th homogeneous subsystem, $\chi_0^{(i)}$ and $j_0^{(i)}$ are the zero-stress susceptibility and magnetization, respectively, of homogeneous subsystem, and $r_x^{(i)}$ and $r_j^{(i)}$ are given by equations 13 and 15, respectively, evaluated using parameters appropriate to each subsystem.

verse problem of determining the structure of a given heterogeneous magnetic system presents great difficulties owing to the non-linearity of the stress dependence in $r_x^{(i)}$ and

materials having the same saturation magnetization and saturation magnetostriction coefficients, at any given state the stress sensitivity is directly proportional to the volume fraction. Note that the higher the stress sensitivity, the greater will be the effect of stress on the magnetic properties of a material, as shown in (18). We also note that, if

$$j_0/j_1 = \chi/\chi_0 = 1/(sS + 1) \quad (22)$$

of the normal deviatoric stress S_1 , then $j_0 = -s$. This can also be written $j_0 = -s j_1$. For randomly directed stresses of magnitude p_i , *Bozorth* gives the initial rate of change of remanent magnetization I_r with applied stress as $j_0 = -I_s/4p_i$, where the numerical factor arises from averaging the internal stresses in all directions, and I_s is the saturation magnetization.

clear that the stress sensitivity, and (18), is not restricted in application to materials in which internal stress determines the magnetic characteristics, although for materials with the reciprocal of the stress sensitivity

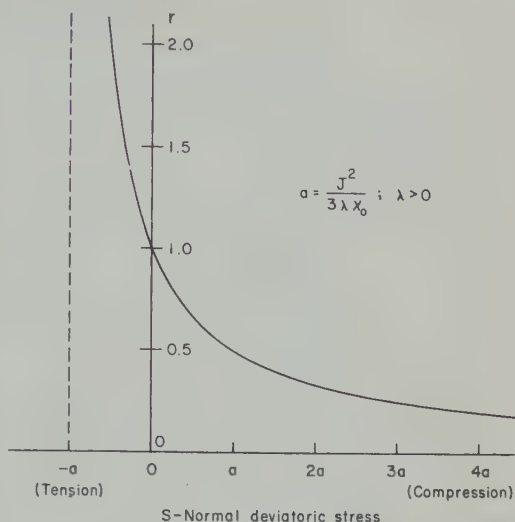


Fig. 2. Ratio r as a function of normal deviatoric stress for positive magnetostriction.

may be thought of in that role. We can thus define a virtual initial stress, a , as

$$a = 1/s = J^2/3\lambda\chi_0 \quad (23)$$

The behavior of magnetic material with a positive magnetostriction coefficient in compression is given directly by the above formulations. Figure 2 gives the ratio r as a function of applied normal deviatoric stress for such a material. The form of the curve is that of an equilateral hyperbola with axial asymptotes $S_1 = -a$ and $r = 0$. The significance of the virtual initial stress, a , as defined by (23) is immediately apparent.

Magnetic material with negative magnetostriction coefficient in compression can be dealt with in the same manner. The resulting formulations will apply over the lower range of applied normal deviatoric stress, but a form of saturation takes place which limits the magnetization that can be acquired along a given axis. When all the available domain magnetizations have become parallel to the stress axis, the material cannot react with a change in magnetization to increased stress. Figure 3 shows the form of the variation of r with applied stress for such a material. Relations of this kind have been reported for susceptibility of rocks by *Kalashnikov and Kapitsa* [1952].

The behavior of materials with positive magnetostriction coefficients in tension is directly

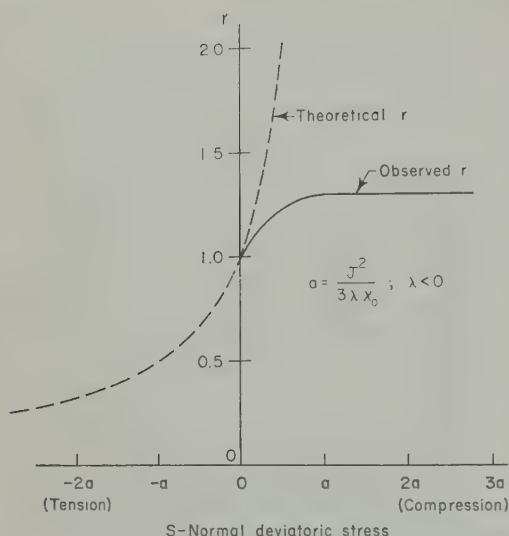


Fig. 3. Ratio r as a function of normal deviatoric stress for negative magnetostriction.

analogous to that of materials with negative magnetostriction coefficients in compression, and vice versa.

In the behavior of real materials, allowance must be made for interaction between magnetic constituents and for closure of magnetic lines within the magnetic system. Such interactions as take place between intimately associated magnetic materials are beyond the scope of the present treatment. It is to be expected that the behavior of rock specimens such as those dealt with by *Graham, Buddington, and Balsley* [1957], would be quite complicated and that analysis of their behavior according to the contributions of constituents would be very difficult, if not impossible.

Test of predicted normal deviatoric stress effects. *Grabovsky and Parkhomenko* [1953] report a series of experiments in which uniaxial stress was applied to magnetite and 'titanomagnetite' ore specimens and the magnetic behavior measured as a function of applied stress. The compositions of the specimens are given in Table 1. Magnetic properties of these specimens are shown in Table 2. In the following sections, the experimental results of *Grabovsky and Parkhomenko* are analyzed with the aim of comparing their results with those predicted from the theory developed in the preceding sections of this article.

In *Grabovsky and Parkhomenko's* experi-

mental configuration (Fig. 4), the specimen was constrained by a brass cylinder with an inner radius of 7 mm and an outer radius of 20 mm. The specimen was seated in this cylinder with low-melting-point Wood's alloy, establishing firm contact with the cylinder walls. This arrangement does not permit the assumption of rigid constraint as used by *Grabovsky and Parkhomenko*. We require the value of the normal component of the deviatoric stress to be parallel to the axis of the applied total stress.

The plane stress solution to the problem of the thick-walled tube under internal pressure is given by *Love* [1927, p. 144]. The radial displacement of the inner wall of the tube is given by

$$U = \frac{p_1 r_1^3}{r_0^2 - r_1^2} \left[\frac{(1 + \nu_2)(1 - 2\nu_2)}{E_2} + \frac{2\nu_2^2}{E_2} + \frac{(1 + \nu_2)}{E_2} \frac{r_0}{r_1} \right]$$

where r_1 = inner radius of tube, r_0 = outer radius of tube, p_1 = internal pressure, E_2 =

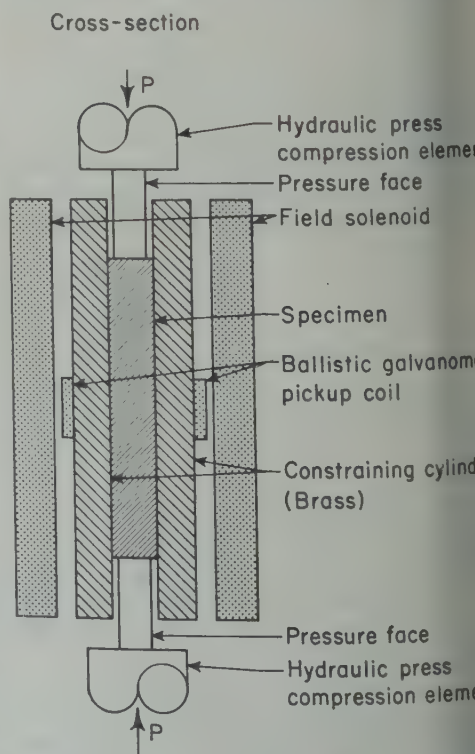


Fig. 4. *Grabovsky and Parkhomenko's* experimental configuration.

TABLE 1. Composition of Grabovsky and Parkhomenko's Specimens

Specimen	Description	Constituent Minerals (Volume Percentage)*			
		Magnetite	Ilmenite	Other Minerals	
				Name	Percentage
	'Cataclysed magnetite ore'	75-80		Calcite	20-25
				Chalcopyrite	Trace
Magnetite	'Ore Mineral Group'	$\frac{1}{2}$ (80-85)	$\frac{1}{2}$ (80-85)	Biotite	15-20
				Clinoclase	
				Carbonate	
				Chalcopyrite	Trace
				Pyrite	Trace

* Approximate volume percentages from *Grabovsky and Parkhomenko* [1953].

modulus of tube, and ν_2 = Poisson's ratio of tube.

inner wall displacements are specified in terms of the deformation of the specimen, they can be expressed as $U = r_1 \sigma_1 \nu_1 / E_1$, where ν_1 = Poisson's ratio of specimen, E_1 = Young's modulus of specimen, and σ_1 = total normal stress in specimen along stress axis. The deviatoric component parallel to the stress axis is $(\sigma_1/3)(\sigma_1 - p_1)$, where σ_1 is the applied stress and p_1 is the internal pressure.

For $\nu_1 = 1/4$, $\nu_2 = 1/3$, $r_1 = 7$ mm, $r_2 = 1$ mm, $E_1 = 6.9 \times 10^{10}$ bars, and $E_2 = 10.3 \times 10^{10}$ bars, we obtain very nearly $S_1 = \sigma_1/2$. Now, using the data of Grabovsky and Parkhomenko and having already been corrected in accordance with the thin wall approximation, their values are

$p_0 = 2\sigma_1/3$, where p_0 is called the uniaxial compression. It is easily seen that $S_1 = 3p_0/4$; this is the correction to be applied to the stresses given by Grabovsky and Parkhomenko. Friction between the specimen and the constraining cylinder have been neglected here.

The stress sensitivities for the materials used in Grabovsky and Parkhomenko's experiments can be calculated directly from their data. Since the massive specimens of magnetite and 'titanomagnetite' ore used in the experiments are not pure, a volume correction is necessary to adjust the behavior of the 'diluted' magnetic material to that of pure material assumed in the development of the theory. The required corrections normalize the specimen parameters to 100 per cent concentration of magnetic material.

TABLE 2. Grabovsky and Parkhomenko's Magnetic Data and Calculated Parameters

Specimen	Uncorrected Data					Calculated Parameters	
	Applied Field H , oersted	Saturation Magnetization j_s' , gauss	Saturation Magnetostriiction, λ'	Initial Susceptibility χ_0' , gauss/oersted	Volume Fraction, v'	Stress Sensitivity s , cm ² /kg	Virtual Initial Stress a , k bars
Magnetite	0	318	3.21×10^{-6}	0.60	0.75	4.76×10^{-4}	2100
	12.5			0.80	to 0.80	6.35	1580
	23.5			0.96		7.62	1310
	0	238	1.06	0.73	0.80	3.94	2540
	3			0.87	to 0.85	4.70	2130
	7			1.06		5.72	1750
	12.5			1.34		7.23	1380
	23.5			1.65		8.90	1120

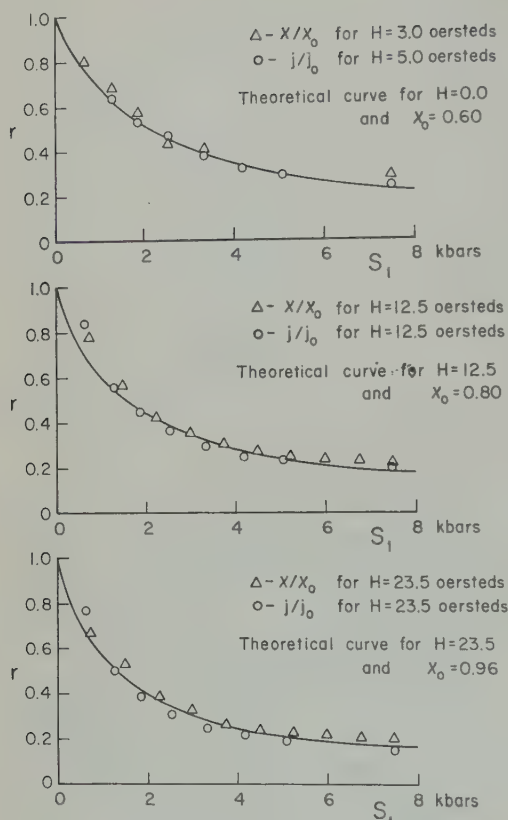


Fig. 5. Comparison of theoretical r with Grabovsky and Parkhomenko's [1953] experimental data for magnetite.

Volume percentages are given by the authors for their specimens. When the volume correction is taken into consideration, (21) assumes the form

$$s = \frac{3\lambda'X_0'}{(j_s')^2} v^{2/3}$$

The calculated stress sensitivities appear in Table 2. Equation 18 is used to calculate the ratios of final to initial susceptibility and magnetization, as a function of applied normal deviatoric stress. These theoretical curves appear in Figures 5 and 6 together with values calculated from the data of Grabovsky and Parkhomenko.

The agreement between the theory and experiment is quite good for the magnetite specimen and fairly good for the titanomagnetite. As can be seen from the constituents listed for the two specimens in Table 1, the magnetite specimen is very nearly composed of one magnetic

mineral constituent, and the titanomagnetite has at least two: magnetite and ilmenite. We should expect, therefore, the assumptions of homogeneity and isotropic susceptibility to be less applicable to the titanomagnetite specimen.

The agreement of the theory developed here with the experimental results of Grabovsky and Parkhomenko supports the use of calculated stress sensitivity as a measure of stability of magnetic materials under the action of directional stress. If a component of magnetization is subject to appreciable change under applied stress, the presence of irreversible effects is to be expected. The value of information embodied in such a component of magnetization to investigations in rock magnetism or paleomagnetism is therefore highly questionable.

Dependence of stress effects on temperature
In the preceding theoretical development, the

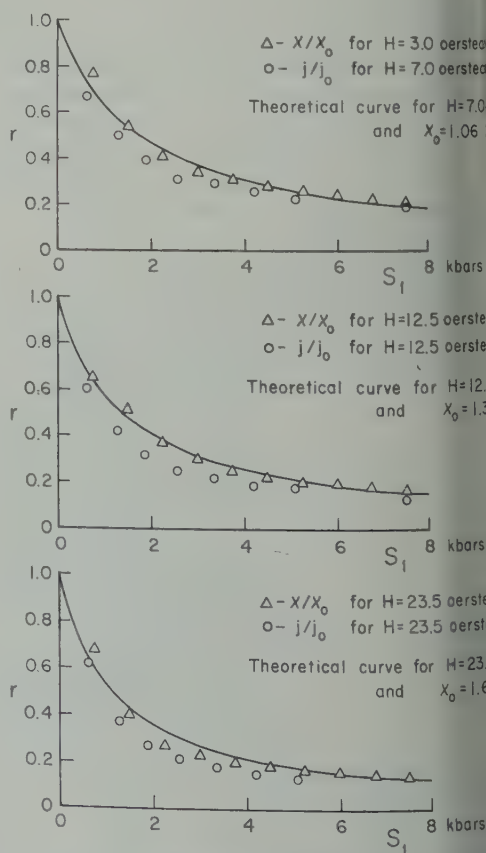


Fig. 6. Comparison of theoretical r with Grabovsky and Parkhomenko's [1953] experimental data for titanomagnetite.

TABLE 3. Data of Kneller and Calculated Virtual Initial Stresses as a Function of Temperature for Nickel

Temperature, °C	0	40	80	120	160	200	240	280	320	
	0.960	0.947	0.936	0.924	0.912	0.898	0.883	0.869	0.852	
$10^{-6}\dagger$	38.5	36.5	34.0	30.4	26.5	22.0	17.0	11.0	5.9	
Sample										
I	28.0	24.0	21.0	18.5	16.0	12.5	9.5	6.5	3.0	
II	24.0	20.0	17.5	15.5	13.5	10.5	8.5	5.5	2.5	
III	18.0	15.0	13.0	11.0	10.0	7.5	5.5	4.0	2.0	
IV	16.0	13.0	11.0	9.5	8.0	6.0	4.5	3.0	1.5	
VII				0.58	0.43	0.36	0.26	0.18	0.08	
IX _s	0.76	0.60	0.47	0.32	0.20	0.14	0.12	0.09	0.04	
X	0.67	0.50	0.40	0.26	0.16	0.09	0.09	0.06	0.02	
Sample										Internal Stress‡
I	2330	2080	1930	1880	1830	1700	1650	1710	1450	2220
II	2000	1730	1610	1570	1550	1430	1480	1450	1200	1900
III	1500	1300	1190	1120	1150	1020	950	1050	970	1430
IV	1330	1130	1010	960	920	820	780	790	730	1300
VII				59	49	49	45	47	39	50
IX _s	63	52	43	32	23	19	21	24	19	
X	56	43	37	26	18	12	16	16	10	

‡ Taken from Table 2, page 85, *Kneller* [1956].

† Taken from Figures 26 and 27, page 112, *Kneller* [1956].

‡ Taken from Table 4, page 103, *Kneller* [1956], calculated from variation of susceptibility with field.

Physical properties are used in developing a relation between applied stress and its effects on magnetic material; it is not restricted to a particular temperature. Given the assumptions listed in the general section of this article, the relations developed apply at any temperature.

The application of this theory to the existence of a multidomain structure in a ferromagnetic material is assumed, the behavior of which depends on the temperature-dependent behavior of the parameters involved in the calculation of stress sensitivity. These parameters have been measured for nickel by *Kneller* [1956]. Nickel is very similar in magnetic properties to iron; hence, some inferences can be made from *Kneller's* data relative to the behavior of magnetic minerals.

The existence of a domain structure depends on the value of the magnetocrystalline anisotropy compared with thermal energy, and the structure in a magnetic system is temperature-dependent. Above a given temperature, the magnetocrystalline anisotropy disappears. Below this temperature and the Curie point, it is found that the type of magnetic behavior predicted by the model of *Verhoogen* [1959] is valid. That is, for this range the mag-

netic behavior would be that of large single-domain grains and the remanent magnetization would reside in regions of crystal dislocations. Elastic strains of the same order of magnitude as those around a dislocation would be required to alter the remanent magnetization. The mechanism described by *Verhoogen* would exhibit the smallest possible stress sensitivity for a given magnetic system. Plastic or slip dislocation deformation of the crystal would be required to affect directions of remanent magnetization carried by such strained regions.

Consider the behavior of the stress sensitivity as a function of temperature. The data of *Kneller* make possible the calculation of this function for nickel. Table 3 gives data taken directly from *Kneller's* curves for seven samples of nickel, together with calculated values of virtual initial stress, a , for different values of temperature. Values of internal stress calculated by *Kneller* for four of the samples from susceptibility variation with applied field are shown on Table 3 for comparison. Internal stress can be obtained only from the magnetic data on the basis of some assumed model. Virtual initial stresses, calculated from (23), are shown plotted as a function of temperature in Figure 7. It is to be

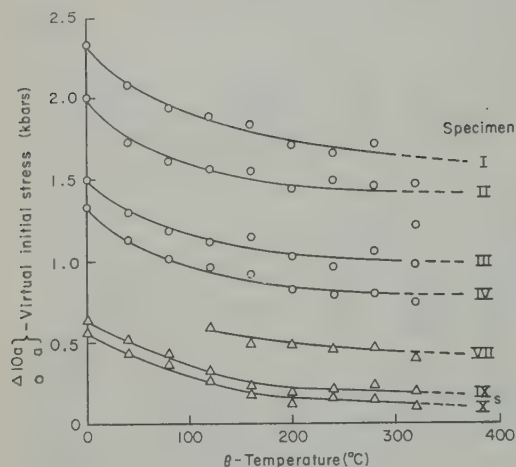


Fig. 7. Virtual initial stress, σ_a , as a function of temperature for nickel, from Kneller's [1956] data.

noted that they approach limiting values as the temperature approaches the Curie point, which for nickel is 360°C . This is significant in that it indicates a finite, nonzero virtual initial stress even at the Curie temperature.

The magnetic properties of nickel are similar to those of magnetite, and the thermoremanent properties of the two materials are similar in many respects. Hence the above result for nickel can be inferred to apply to magnetite at high temperature. Measurements of the saturation magnetization and magnetostriction coefficients and the ferromagnetic susceptibility of magnetite and other rock-forming magnetic minerals as a function of temperature are required to validate this inference. The failure of some experiments to demonstrate significant stress effects on thermoremanent magnetization [Stott and Stacey, 1959, 1960; Kern, 1961] does, however, support such a conclusion. Within the framework of the theory developed here, and from such experimental evidence as is presently available, the conclusion can be reached that stress sensitivity can be used to describe the stress-dependent magnetic behavior of ferromagnetic minerals at any temperature.

Acknowledgments. This work was done at the University of California, Berkeley, California, aided by grants from the American Petroleum Institute and the Institute of Geophysics of the University of California, and was partially carried out under a Fellowship of the American Chemical Society Petroleum Research Fund.

REFERENCES

- Becker, R., Elastic strains and magnetic properties, *Physik. Z.*, **33**, 905-913, 1932.
- Becker, R., The technical magnetization of iron, *Magnetismus*, Debye, Leipzig, 82-90, 1933.
- Bloch, F., Theory of the exchange problem and residual ferromagnetism, *Z. Physik*, **76**, 513-527, 1932.
- Bozorth, R. M., *Ferromagnetism*, D. Van Nostrand, New York, 968 pp., 1951.
- Grabovsky, M. A., and E. I. Parkhomenko, Changes in the magnetic properties of magnetite under high compressive pressures, *Izvest. Akad. Nauk SSSR Ser. Geofiz.*, **6**, 405-417, 1953.
- Graham, J. W., A. F. Buddington, and J. L. Balsley, Stress-induced magnetizations of sedimentary rocks with analyzed magnetic minerals, *Geophys. Research*, **62**(3), 465-474, 1957.
- Graham, J. W., A. F. Buddington, and J. L. Balsley, Magnetostriction and paleomagnetism of igneous rocks, *Nature*, **183**, 1318, 1959.
- Jeffreys, H., *Cartesian Tensors*, Cambridge University Press, London, 1931.
- Kalashnikov, A. G., and S. P. Kapitsa, Magnetic susceptibility of rocks under elastic stresses, *Doklady Akad. Nauk SSSR*, **86**(3), 521-523, 1952.
- Kern, J. W., Effects of moderate stresses on distributions of thermoremanent magnetization, *Geophys. Res.*, **66**, 3817-3820, 1961.
- Kneller, E., Ueber die Temperaturabhängigkeit der Bestimmungsgrossen der technischen Magnetisierungskurve von Nickel, *Beiträge zur Theorie des Ferromagnetismus und der Magnetisierungskurve*, edited by W. Köster, Springer, Berlin, 82-139, 1956.
- Love, A. E. H., *A Treatise on the Mathematical Theory of Elasticity*, Dover, New York, 643 pp., 1944.
- Neel, L., Some theoretical aspects of rock magnetism, *Phil. Mag. Suppl. Advances in Physics*, **4**, 191-243, 1955.
- Sokolnikoff, I. S., *Mathematical Theory of Elasticity*, McGraw-Hill Book Co., New York, 415 pp., 1956.
- Stacey, F. D., Effect of stress on the remanent magnetism of magnetite-bearing rocks, *J. Geophys. Research*, **63**(2), 361-368, 1958.
- Stott, P. M., and F. D. Stacey, Magnetostriction and paleomagnetism of igneous rocks, *Nature*, **183**, 384-385, 1959.
- Stott, P. M., and F. D. Stacey, Magnetostriction and paleomagnetism of igneous rocks, *J. Geophys. Research*, **65**, 2419-2424, 1960.
- Verhoogen, J., The origin of thermoremanent magnetization, *J. Geophys. Research*, **64**(12), 2449, 1959.
- Weiss, P., Hypothesis of the molecular field of ferromagnetic properties, *J. Phys.*, **6**(4), 661-670, 1907.

(Manuscript received May 5, 1961; revised August 18, 1961.)

Stress Stability of Remanent Magnetization

JOHN W. KERN¹

*University of California
Berkeley, California*

Abstract. In this paper a criterion will be developed for determining the stress required to affect a given component of remanent magnetization. This criterion is associated with the process of alternating-field (a-f) demagnetization currently in use for eliminating unstable components of magnetization from paleomagnetic specimens. The conclusion is drawn that if magnetoelastic coupling, or magnetostriction, plays a role in determining directions of remanent magnetization, such effects can be minimized or removed by a-f demagnetization. Further, the reliability for paleomagnetic measurements of components of magnetization with greatest resistance to a-f demagnetization is higher than for those easily demagnetized, especially if stress effects are thought to be present.

Alternating-field demagnetization. The process of a-f demagnetization is carried out in the majority of laboratories as a discontinuous procedure. Measurements of the remanent magnetization of a sample are alternated with applications of a sinusoidally varying magnetic field. The field is slowly increased from zero amplitude to the desired maximum for the particular application, maintained at this maximum for a period of time, and then slowly decreased to zero. The maximum amplitude of the field is increased by discrete amounts for each successive application. Time-dependent effects are sometimes encountered in the measurement of specimens. Such effects can be explained on the basis of mechanisms accounting for so-called viscous magnetization [Doell, 1955].

In the a-f demagnetization of rock specimens used in paleomagnetic research, care is usually taken to apply the field in such a manner that preferred axes are introduced along which components of magnetization might be only partially affected by the applied field. This requires some method of rotating the specimen relative to the applied field at a rate which is compared with the frequency of the applied sinusoidal field and incommensurate with the phenomenon of anhysteretic magnetization (i.e., magnetization arising from the effect of an alternating field biased by a constant field); this effect gives magnetizations much greater than would result from the application

of the biasing field alone) due to the biasing effect of the earth's field is also minimized [Neel, 1955].

Domain wall translation model for alternating-field demagnetization. One model which accounts for demagnetization by the procedure described above is based on Neel's [1955] and Kondorski's [1937] theory of coercive force.

For a homogeneous aggregate containing a large number of domains associated in a multidomain structure, the vector sum of the domain magnetic moments gives the magnetic moment of the aggregate, if closure of magnetic lines within the material is neglected. Application of an alternating magnetic field in the manner outlined in the preceding section results in translations of domain walls across free paths determined by the magnitude of the applied field and the nature of the opposition to wall movement inherent in the material. In general, the greater the applied field, the greater the free path for domain wall translation [cf. Neel, 1955, pp. 228 ff.]. A slow decrease of applied alternating field leads to a decrease in these free paths and finally to fixing of domain walls in positions randomly distributed in space. The resultant magnetization of the volume of material swept by the domain walls thus will be reordered. If the original ordering of the domain wall positions gave rise to a net magnetic moment, the random reordering of these domain magnetizations results in a reduction in their net magnetization and in a random direction in space for their resultant. In general, for populations of domains with significant remanent mo-

¹Now at the Rand Corporation, Santa Monica, California.

ment, the number of domains involved is very large, and hence the resultant will be insignificant compared with the original magnetization.

An a-f demagnetization process, therefore, represents a succession of operations which reorder fractions of the domain population that contribute to the original remanent moment of the specimen.

Microscopic coercive force. The microscopic coercive force for a given region in a magnetic system is defined here as that internal effective field causing displacement of domain walls across the region. Use of the term microscopic distinguishes this definition from those associated with the hysteresis curve for a magnetic material. In the theoretical developments of Neel, Kersten, and others, the term coercive force is used in the sense here denoted by *microscopic coercive force*.

On the basis of the model discussed in the preceding section, the a-f demagnetization curve represents a distribution of microscopic coercive forces for a given magnetic system. The values of the applied field must be corrected in the manner given by Uyeda [1958], then the magnetization remaining after application of succeeding increments of field is plotted as a function of applied field. This a-f demagnetization curve, proceeding from zero to higher applied fields, gives a measure of the relative amounts of magnetization contributed by regions of successively higher microscopic coercive force. In this process, application of a given field to a particular material results in translation of domain walls through regions with microscopic coercive forces less than or equal to the effective internal field.

Thus the domain structure has been further analyzed into regions characterized by their microscopic coercive force. A single domain with a given spatial distribution of obstacles to domain wall movement can be analyzed in this way into regions, or intradomain volumes, characterized by the minimum field required to drive a domain wall across the volume.

In the a-f demagnetization process, the magnetization associated with regions of microscopic coercive force less than the applied field amplitude will be randomly reordered. That is, if a domain wall lies in a region of microscopic coercive force less than the applied field, the

wall will be translated within this region. Random reordering of the wall positions for a number of such regions leads to a random reordering of the portion of specimen magnetization associated with them. This viewpoint is the same as that discussed in the preceding section.

Microscopic coercive force and minimum coercive stress. Consider the portion of magnetization of a remanent magnetic system associated with microscopic coercive forces greater than H_m . This portion will not be affected by fields less than H_m . From the energy function for a system which is homogeneous with respect to volume magnetization given by $dg = e_{ij}dS_{ij} + j_i dH_i$ [Kern, 1961], the following relation can be derived for isotropic magnetostriction:

$$1/j(\partial H/\partial S)_i = 3\lambda/J^2$$

where λ is the saturation magnetostriction, J is the saturation magnetization, j , H , and S are the components of magnetization, field, and normal deviatoric stress, respectively, in the same direction, and the magnetoelastic strain e is given by $3\lambda/2[(j/J)^2 - 1/3]$.

If the portion of the magnetization associated with H_m can be separated into sets with homogeneous magnetization, this relation can be applied to each set. Since the magnetization remains constant for such a set until the field exceeds H_m , the above equation can be integrated for $H \leq H_m$. Thus

$$dH/dS = j3\lambda/J^2$$

$$\int_0^{H_m} dH = 3\lambda j/J^2 \left[\int_0^{S_m} dS_1 \right]$$

and

$$H_m = 3\lambda j S_m / J^2$$

Hence,

$$S_m = J^2 H_m / 3\lambda j$$

where the stress S_m is that stress which will cancel the effect of H_m on the magnetization for a positive magnetostriction coefficient.

We note that any H less than H_m has no effect on j . Hence any stress less than S_m will have no effect on j ; i.e., S_m is the minimum stress affecting j .

To get a lower bound to S_m , we can let j assume its maximum value J . Thus the minimum

fecting regions of microscopic coercive force H_m will be

$$S_m = H_m J / 3\lambda$$

ue of stress will be used to define *minimum coercive stress*.

magnetite, $J = 450$ gauss and $\lambda = 10^{-6}$ [Weisz, 1954], and hence $S_m = 3.8H_m$ is obtained. Thus a component of magnetization associated with coercive forces of 500 oersteds will require a normal deviatoric stress of 1900 dynes/cm² to effect a change in its magnitude. *Minimum coercive stress and the alternating-magnetization curve.* The correspondence between microscopic coercive force and what has been termed the minimum coercive stress is shown in the preceding section. It has been noted that the a-f demagnetization curve gives a representation of the distribution of microscopic coercive forces for a given remanent magnetic system. It remains to be shown that a spectral representation of these microscopic coercive forces can be obtained and that such a representation serves as a spectral representation of the minimum coercive stresses of the system. The portion of the remanent magnetization curve for a particular increment of applied field $\delta H(H_m)$ is given by

$$\left[\frac{\delta j}{\delta H} (H_m) \right] \delta H, \text{ where } \left[\frac{\delta j}{\delta H} (H_m) \right]$$

slope of the a-f demagnetization curve for the interval δH centered on the applied field of magnitude H_m .

Recall that the a-f demagnetization process is remanent magnetization without regard to the direction of original magnetization. The rotation of the specimen in the applied field leads to the alternating field being directed along the directions of magnetization of all regions in the magnetic system. However, we are concerned with the variation of normal deviatoric stress in one direction only. Thus, if an applied stress is increased from zero to a given value, it will affect only the regions with directions of magnetization such that the component of normal deviatoric stress parallel to the magnetization is greater than the minimum coercive stress for those regions.

If S_m is the minimum stress required to effect a change in magnetization for a set of

regions having microscopic coercive force H_m , then the *maximum* variation of magnetization j with S_m is given by

$$\begin{aligned} \left[\frac{\delta j}{\delta S} (S_m) \right] &= \left[\frac{\delta j}{\delta H} (H_m) \right] \left(\frac{\partial H}{\partial S} \right)_{j=S} \\ &= \frac{3\lambda}{J} \left[\frac{\delta j}{\delta H} (H_m) \right] \end{aligned}$$

where $[(\delta j / \delta H)(H_m)]$ is obtained, as noted above, from the a-f demagnetization curve. The actual variation of remanent magnetization with applied stress is therefore given by the inequality

$$\frac{\delta j}{\delta S} < \frac{3\lambda}{J} \left[\frac{\delta j}{\delta H} (H_m) \right]$$

Owing to the variation of stability of components of magnetization with different microscopic coercive forces, the directions of remanent magnetization associated with high microscopic coercive forces are more reliable for paleomagnetism and rock magnetism research. As noted by Bol'shakov [1957] and Uyeda [1958] this portion of the magnetization corresponds to high-temperature thermoremanent magnetization in igneous rocks. Crystallization magnetization [Doell, 1956; Kobayashi, 1959; Haigh, 1958] provides a mechanism of high-stability remanent magnetization in sedimentary rocks. Highly stable remanent magnetization is less affected by applied stress, and is in fact not affected by stresses less than the above-defined minimum coercive stress. The a-f demagnetization process, therefore, provides a means of minimizing the use of erroneous directions of magnetization (i.e., those not representing directions of ambient fields) in paleomagnetic investigations.

Acknowledgments. This work was done at the University of California, Berkeley, California, aided by grants from the American Petroleum Institute and the Institute of Geophysics of the University of California, and was partially carried out under a Fellowship of the American Chemical Society Petroleum Research Fund.

REFERENCES

- Bol'shakov, A. S., The stability of normal rock magnetization, *Izvest. Akad. Nauk SSSR, Ser. Geofiz.*, 5, 53-62, 1957.
Doell, R. R., Remanent magnetism in sediments, Doctoral dissertation, University of California, 156 pp., 1955.

- Doell, R. R., Remanent magnetization of the Upper Miocene 'blue' sandstones of California, *Trans. Am. Geophys. Union*, 37, 156-167, 1956.
- Haigh, G., The process of magnetization by chemical change, *Phil. Mag.*, 3(27), 267, 1958.
- Kern, J. W., The effect of stress on the susceptibility and magnetization of a partially magnetized multidomain system, *J. Geophys. Research*, 66(11), 3807-3816, 1961.
- Kobayashi, K., Chemical remanent magnetization of ferromagnetic minerals and its application to rock magnetism, *J. Geomag. Geoelec.*, 10(3), 99, 1959.
- Kondorski, E., H_c and irreversible changes in magnetization, *Physik. Z. Sowjetunion*, 11, 58, 1937.
- Neel, L., Some theoretical aspects of rock magnetism, *Phil. Mag. Suppl. Advances in Physics*, 4, 191-243, 1955.
- Uyeda, S., TRM as a medium of paleomagnetism with special reference to reverse TRM, *J. Geophys.*, 2(1), 1-123, 1958.
- Weisz, R. W., Magnetic anisotropy constant of ferromagnetic spinels, *Phys. Rev.*, 96, 80, 1954.

(Manuscript received May 5, 1961; revised August 18, 1961.)

Observations of Internal Waves near Hudson Canyon¹

ROY D. GAUL²

U. S. Navy Hydrographic Office
Washington, D. C.

Abstract. Multi-level water temperature measurements made during the spring and fall at stations located in a small triangular array indicated pronounced activity of internal waves having nominal periods of several minutes. Individual wave velocities were determined and found to fall within similar narrow ranges of speed and direction during the observation periods in both seasons. These results indicated that the waves typically travel shoreward at speeds of 0.75 to 1.3 knots in the observation depth of 185 feet.

INTRODUCTION

The existence of internal waves beneath the surface has long been known, but only recently has observation technology advanced to the point of making controlled detection practical. *Keulegan* [1904] explained 'dead water' in terms of internal waves generated by a vessel moving slowly through a shallow, low-density stratified layer. The free wave speeds that he calculated (which happen to correspond closely to those cited in this report) could not account for propagation speeds of internal waves of this character until *Haurwitz* [1950] and *Long* [1950] took into consideration the effect of Earth rotation. In the interim, *Fjeldstad* [1950] had theoretically examined internal waves within a continuously varying density stratification as opposed to the original two-layer model. These works and others too numerous to list [see *Davis and Patterson*, 1956] were concerned chiefly with oscillatory waves of relatively long period. The work of *Keulegan* [1953] and *Long* [1956] dealt specifically with the propagation of a water internal solitary wave in a two-layer medium. Their theoretical studies are of great interest because much of the data obtained during the course of the investigations reported in this paper clearly indicated the presence of internal perturbations bearing little resemblance to harmonic or random variations

and having discrete features identifiable in groups of one to four waves.

FIELD MEASUREMENTS

Field data for this study were collected during the fall of 1959 and the spring of 1960 in the vicinity of Texas Tower 4. The tower (hereafter designated TT4) was located about 70 miles southeast of New York Harbor on the north shoulder of the Hudson Canyon (Fig. 1) in water 185 feet deep.

The sensing system is shown schematically in Figure 2. A spherical, Japanese glass fishing float, 14 inches in diameter, provided buoyancy, and the net woven around the float was coupled to a 3/32-inch steel cable which was linked to a 100- to 150-pound gravity anchor. The 3/8-inch electrical lead cables from each resistance thermometer were taped or tied to the 3/32-inch cable.

The array used during the fall of 1959 consisted of two buoys 500 feet apart which formed an equilateral triangle with the vertical guide cables below the hydrographic room of the tower (Fig. 3). The two submerged buoys on the east-west base of the 500-foot triangle supported resistance thermometers at nominal depths of 65 and 100 feet below mean water level (MWL). Thermometers at the tower were at 65, 80, 100, and 130 feet below MWL. During the spring of 1960 the array consisted of three buoys positioned as shown in Figure 4 with resistance thermometers at approximately 50, 70, 90, and 110 feet below MWL. Note that the equilateral triangle was 400 feet on a side instead of 500 feet. To minimize the measure-

¹Contribution from the Department of Oceanography and Meteorology of the Agricultural and Mechanical College of Texas.

²Presently at the Department of Oceanography and Meteorology of the Agricultural and Mechanical College of Texas, College Station, Texas.

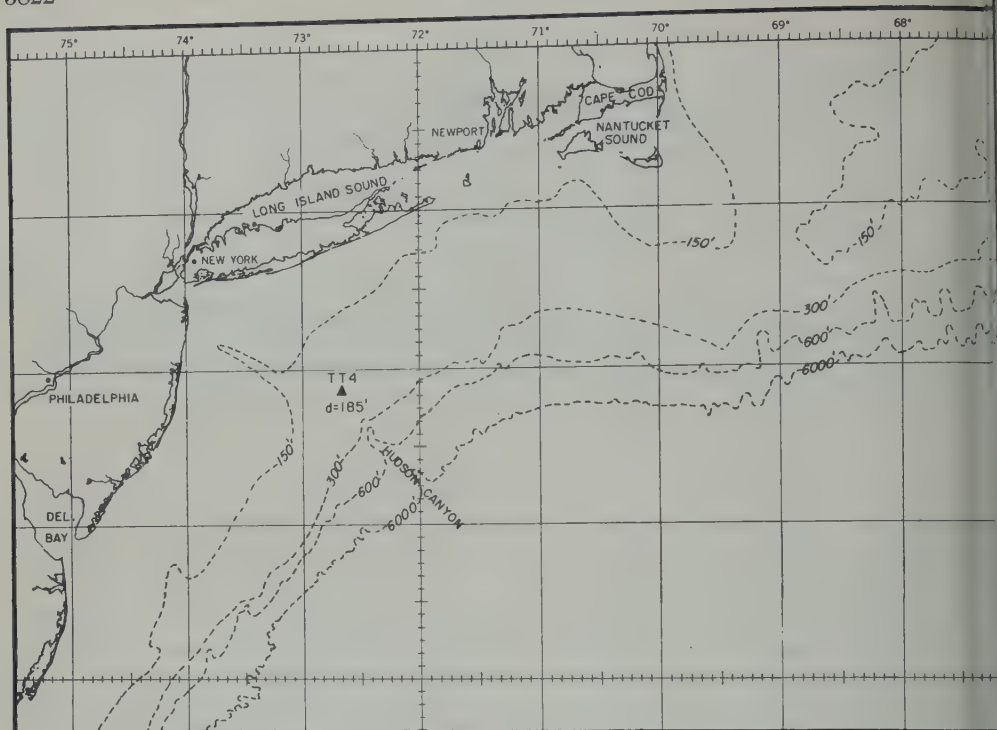


Fig. 1. Location of Texas Tower 4 near Hudson Canyon.

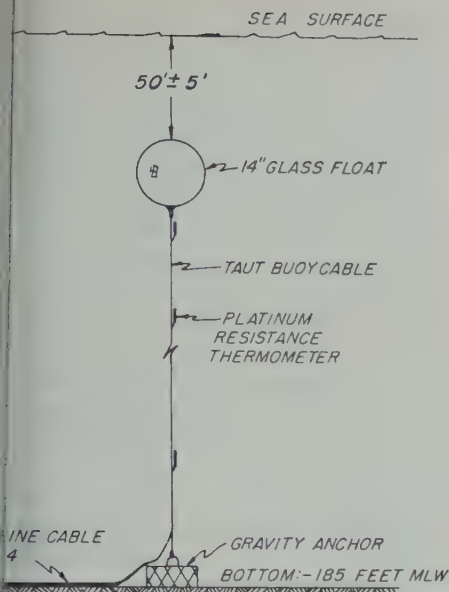
ment of turbulent fluctuations caused by water flow past the tower substructure, the nearest buoy in the array was located 250 feet south of the tower.

The buoy arrays were installed by lowering them from a 26-foot surf boat stationed on the tower. In the fall, sextant angles taken from the boat and supplemented by sight lines relative to tower caissons were used to position the two buoys. In spite of drift rate, sea conditions, sextant and sighting errors, and the 40- to 50-foot free fall of the buoy when cut loose over station, the actual positions of the buoys are estimated to be within 40 feet of the predetermined position. Considerably improved positioning techniques were used in the spring of 1960. Azimuth instruments were positioned at the end points of an accurately measured base line on the tower deck. Sighting angles were preset to establish an intersect at a given buoy position with a precision better than 1 foot. The boat was maneuvered into position according to flag signals given by the men stationed at the azimuth instruments. When the line restraining the partially submerged buoy was

centered in the sights of both instruments was cut and the single station array permitted to fall freely into place. The locations of buoys are estimated to be within 10 feet of calculated station positions shown in Figure 1.

Two 4-channel Minneapolis-Honeywell 'B' balancing potentiometer recorders were used during the fall to record resistance thermometer output on a 0° to 30°C full scale. The sampling rate for a given channel depended on the temperature differences between the channels and normally averaged 6 to 10 seconds. For the spring survey the outputs from four thermometer levels of all three buoy stations were recorded on a single recorder modified to handle 12 different information channels. The sampling interval for an individual information channel was usually from 30 seconds. GMT time marks were manually scribed on the records at 1- to 8-hour intervals. Precision of these marks is estimated to be 1 second for the majority of chart speeds.

The selection of resistance thermometer temperature sensing was largely a matter of convenience. Both thermometers and re-



Schematic elevation of submerged buoy station.

readily available and the normalized reading on a scaled strip chart promised a reduced data-processing effort. The main objective of the type of resistance thermometer used in this experiment was its long time constant compared with other temperature sensors currently on the market. From fairly crude step response tests of the entire measuring system (thermometers and recorders) it is estimated that the time constant for the system was in the neighborhood of 60 seconds.

ANALYSIS AND RESULTS

The data analysis technique consisted essentially of subjectively selecting cases from the temperature records wherein identifiable wave signatures, i.e., distinctive temperature variations related in form (but not necessarily in magnitude), could be ascertained to have occurred in a single horizontal plane at all three measuring stations. The times of passage at each station were then used to calculate the speed and direction of the signatures, these signatures assumed to have a direct relationship to internal waves. Ninety-five cases were recorded during the fall period (September 28 to October 15, 1959) and 201 cases from the spring period (May 13 to June 4, 1960). An attempt to preserve some measure of

statistical significance all records were examined for each data period, and whenever the wave signature relative to background temperature variations was distinctive enough so that its passage at each station could be unmistakably recognized the data were used. When a 'train' of several wave signatures occurred, the most prominent one was selected, or several were averaged. The magnitude of temperature variation was not considered; distinctive features and uniform shape of the signature observed at each station were the only criteria used for selection. The middle of the temperature peak (internal wave trough) was normally used to index the time of passage at each station.

No attempt was made to ascertain or correct for the horizontal movement of the water medium in which the waves were traveling. Accurate ocean current observations were not taken concurrently. During one period of two weeks, however, reliable hourly current measurements were taken at 100 feet below MLW. Grouping of these data into the four direction quadrants resulted in the following averages for current

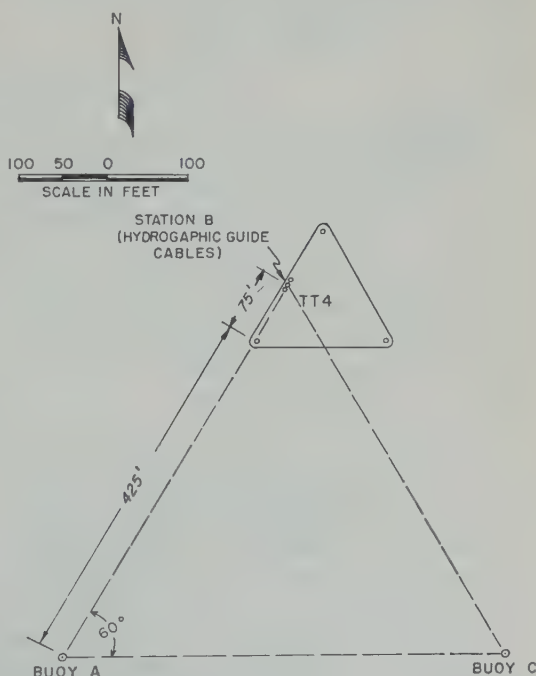


Fig. 3. Plan layout of submerged temperature-measuring stations, September-October 1959. Resistance thermometers were at 65 and 100 feet below the water surface at each buoy.

ility of internal solitary wave theory
d by Long [1956] and Keulegan [1953].
of these theoretical treatments a two-
system is assumed, and Long further
a rigid boundary at the water surface
as at the bottom. Keulegan's work was
e selected as being most appropriate.

His equations are summarized below and then
applied to several observed cases of tempera-
ture-salinity structures for comparison with
measured wave signature speeds.

The mathematical model consists of a two-
layer fluid medium of thickness h_1 and density
 ρ_1 between the free surface and the internal

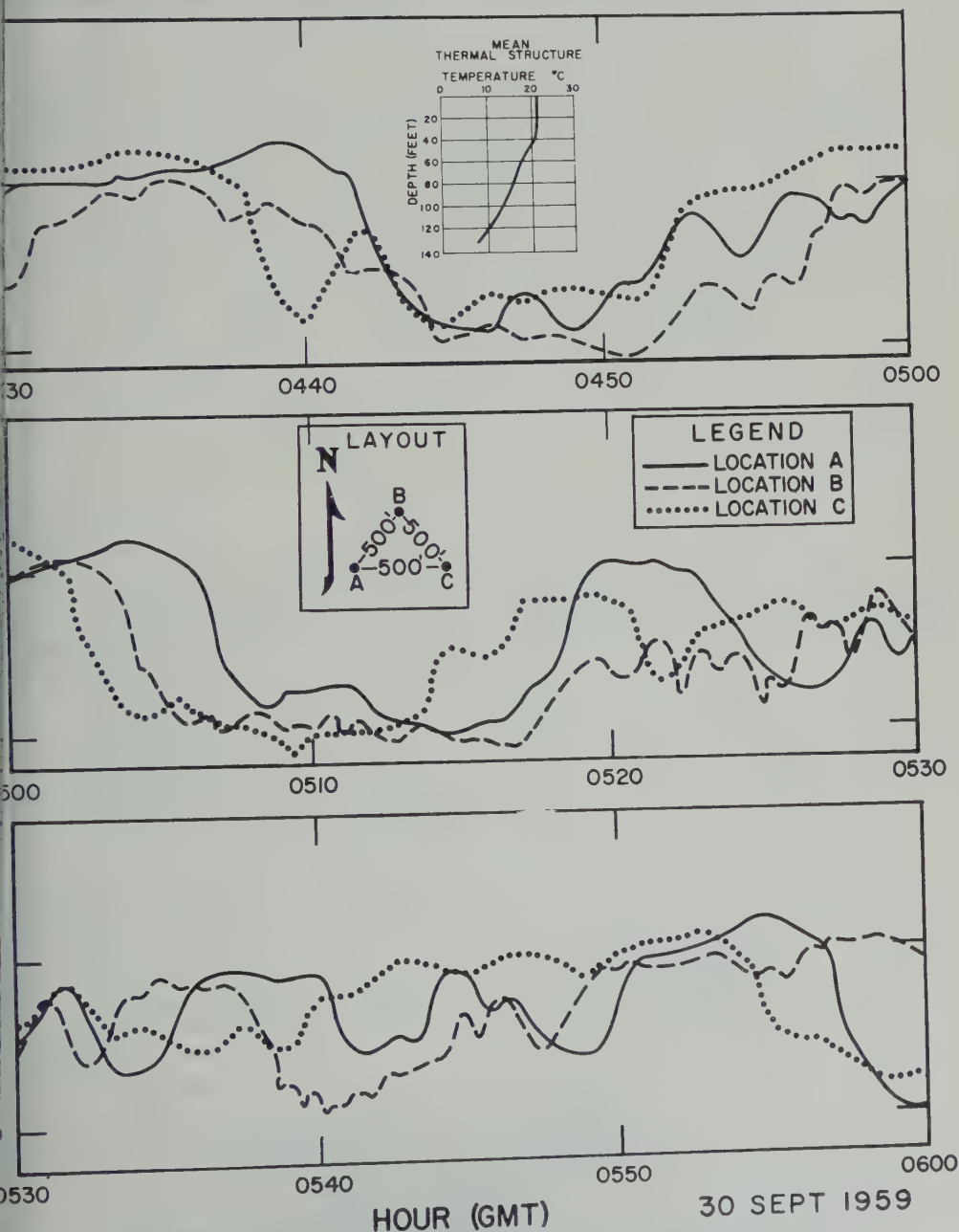


Fig. 5. Sample of three-station temperature recording at 100 feet below the water surface.

density discontinuity overlying a layer of thickness h_2 and density ρ_2 between the discontinuity and a rigid bottom. It is shown that a solitary wave of infinitesimal height occurring at the internal discontinuity will travel at a phase speed c_0 , given by

$$c_0 = \left[\frac{gh_1h_2}{h_1 + h_2} \left(\frac{\rho_2 - \rho_1}{\rho_2} \right) \right]^{1/2} \quad (1)$$

where g is acceleration due to gravity. The equation is the same as that given by De [1950] and Haurwitz [1950] for the oscillation of an internal long wave, neglecting the effect of Earth rotation. Ekman [1904] also applied this equation to explain the 'dead water' effect on vessels moving slowly in a shallow low-density surface layer. The celerity c of the internal wave of finite amplitude η_0 is shown to be

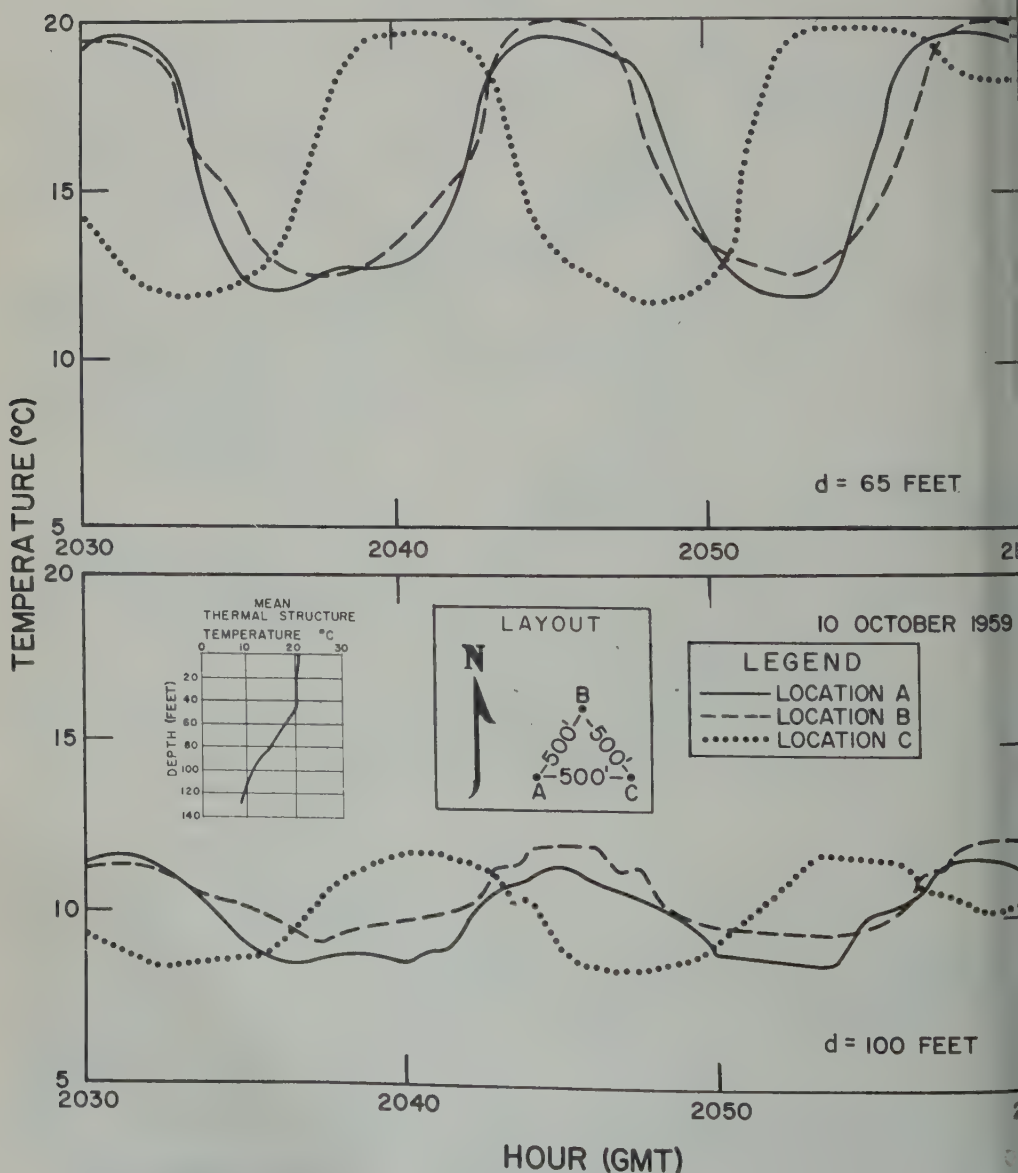


Fig. 6. Two-level temperature recordings at three stations indicating the passage of an internal wave train.

$$c = c_0 \left[1 + \left(\frac{h_1 - h_2}{h_1 h_2} \right) \eta_0 \right]^{1/2} \quad (2)$$

η_0 is relative to the undisturbed level of salinity discontinuity. From (2) it is seen when the discontinuity is at mid-depth the speed is independent of internal wave depth.

A precise comparison of individual signature speeds observed at TT4 to Keulegan's theory is not possible because salinity was not measured continuously and at the same depths as temperature. Water samples for salinity titration were taken at several depths daily from May 13 to June 3, 1960. The observed average salinities varied from 31.4 per mil at the sur-

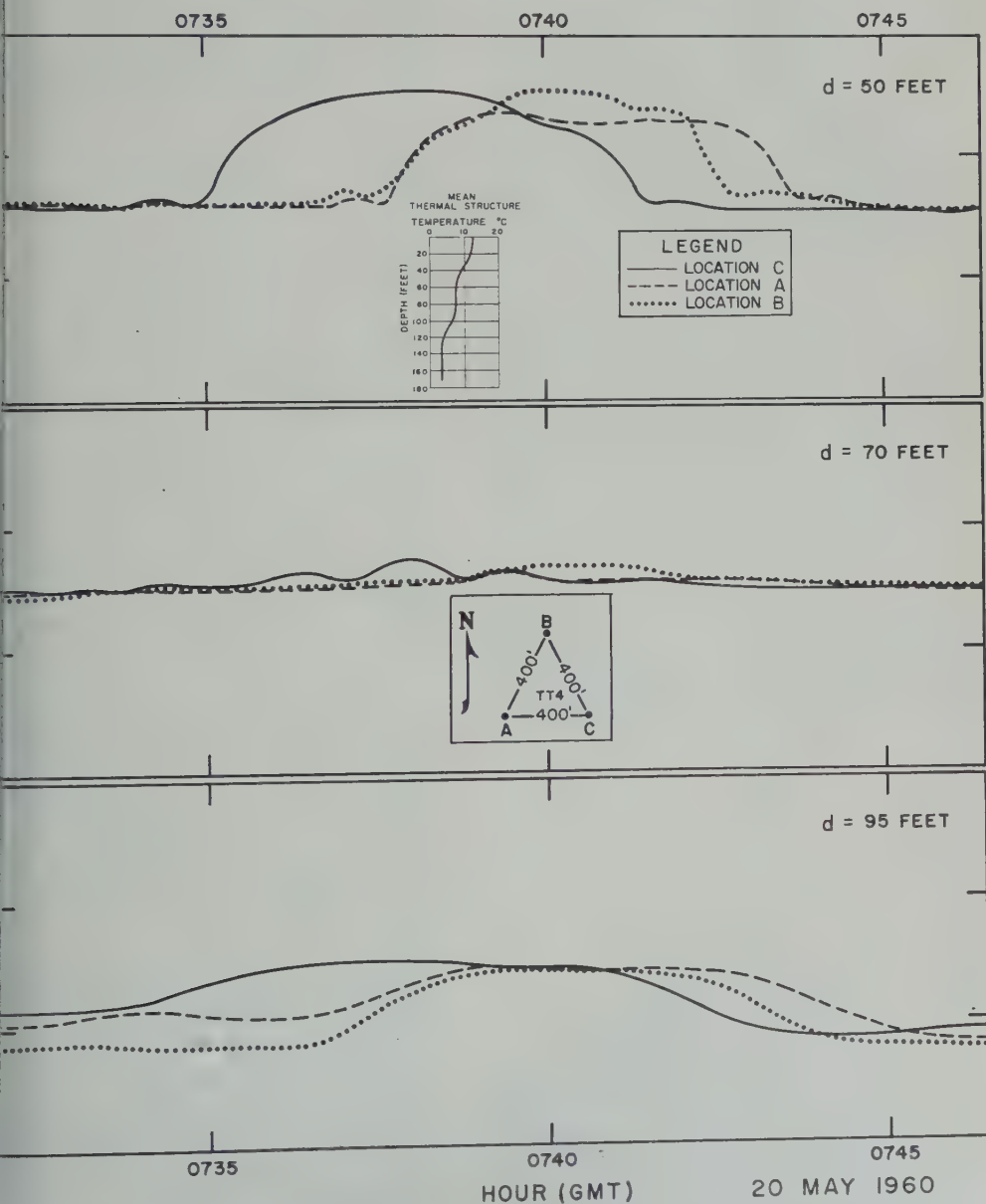


Fig. 7. Three-level temperature recordings at three stations indicating the passage of an internal solitary wave during existence of a double thermocline.

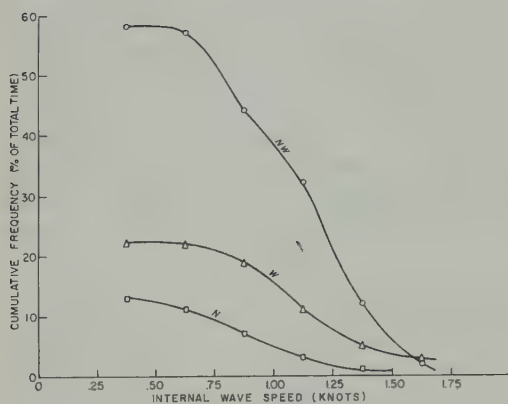


Fig. 8. Cumulative frequencies of occurrence of internal waves observed in the fall of 1959 (based on 95 cases selected from the period September 28-October 15). Directions shown are of wave progress within 45° sectors.

face to about 32.5 per mil at and below 90 feet depth. For purposes of two-layer calculation of wave speeds during the spring period, values of 31.5 per mil for salinity S_1 in the upper layer and 32.5 per mil for salinity S_2 in the lower layer have been assumed. Salinity was assumed to be 32.0 per mil in both layers during the observation period in the fall of 1959.

The average speed of the wave signatures shown in Figure 6 is 0.95 knot. To apply (1) and (2), the mean thermal structure shown in the inset of Figure 6 was replaced by a sharp discontinuity at 70 feet. Temperature in the upper layer was taken to be 20°C , and in the

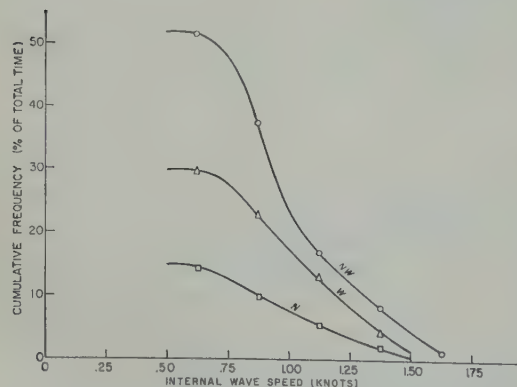


Fig. 9. Cumulative frequencies of occurrence of internal waves observed in the spring of 1960 (based on 201 cases selected from the period May 13-June 4). Directions shown are of wave progress within 45° sectors.

lower layer 10°C . The resulting densities $\rho_1 = 1.0225 \text{ g/cm}^3$ and $\rho_2 = 1.0247 \text{ g/cm}^3$. The mean observed temperature gradient at 650 was about 0.25°C per foot, so the temperature change of 7°C at that level indicates that the actual internal wave height η_0 was 28 feet. Substituting these values into (1) and (2) gives a value of 0.96 knot for the theoretical speed. The extremely good agreement between

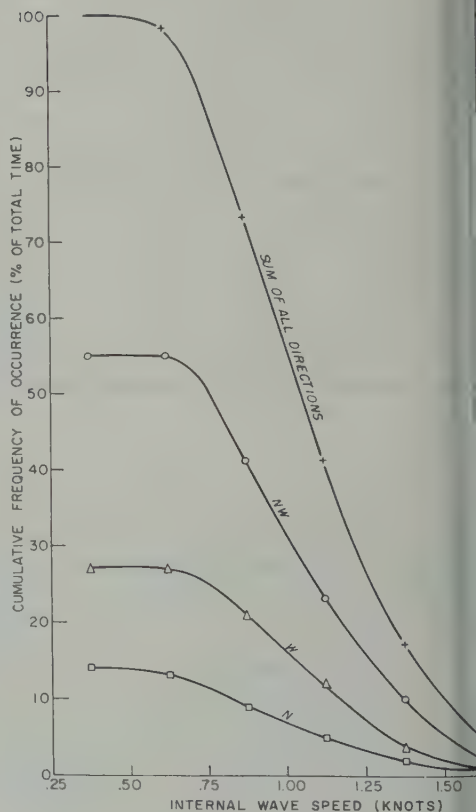


Fig. 10. Cumulative frequencies of occurrence of all internal waves observed (based on 201 cases). Directions shown are of wave progress within 45° sectors.

observed and calculated speeds must be considered largely coincidental in view of the numerous theoretical simplifications and the errors inherent in the actual measuring system.

Further direct comparisons were drawn from five other cases, selected at random, for which reasonable estimates of mean temperature and salinity distributions could be made. Assumptions similar to those cited above were made to reduce the observed temperature-salinity

TABLE 1. Comparisons of Observed and Theoretical Internal Solitary Wave Speeds

Time (GMT), h m	T_1 , °C	T_2 , °C	η_0 , ft	Ob- served Speed, knots	Theo- retical Speed, knots
59	20 40	20 10	28	0.95	0.96
60	9 55	10 4	25	0.65	0.85
60	7 40	12 8	20	0.45	1.25
60	21 50	12 5	40	0.80	1.15
60	22 50	12 5	20	0.88	1.00
60	10 40	12 5	24	0.87	0.90

tion to the two-layer case. Results of the comparisons are given in Table 1.

DISCUSSION AND CONCLUSIONS

The study described in this paper is limited to a small area in a particular location having a depth of 185 feet. Nevertheless, studies made there, some of which have been cited, strongly indicate that waves of the type observed at TT4 occur frequently in shallow water. It is probable that the lack of data on these waves is more to the paucity of experiments than to the nonexistence or restricted occurrence of the phenomena. As instrumentation and data techniques are improved, continuous time series observations may well reveal quite a broad spectrum of significant, and heretofore unobserved, internal activity in the subtidal frequency range.

Comparison of individual wave forms recorded at each of the three stations at TT4 raises the question of uniformity. Temperature amplitudes of half-wavelengths were not usually identical. Three possible explanations for this are: (a) the internal wave crests were closer than the distance across the detection area; (b) the temperature sensing elements did not lie at the same elevation at each station; (c) some of the individual internal wave forms were sloped relative to the bottom. It is certainly unlikely that the prefabricated buoys were exactly uniform or the bottom perfectly flat. Vertical tolerances of 2 feet in sensing depths are quite reasonable. Any error of this order other than current-induced tilt of the wire buoy, would cause a consistent amplitude deviation at a particular station.

The particular case of internal wave passage

illustrated in Figure 8 is worthy of special notice. In this case two thermoclines were separated by an isothermal layer. It so happened that the resistance thermometers were distributed so that one was in each of the layers both before and during the wave passage. From the resulting record it is obvious that both thermoclines moved downward and back to their initial elevations in unison. The wave form is distinctly solitary, and the difference, if any, in wave amplitudes at 50 and 95 feet appears to be small. This observed case definitely indicates a single mode of internal oscillation.

The comparison of the internal wave observations to the solitary wave theory of *Keulegan* [1953] must necessarily be crude since the actual thermal structure was usually characterized by a thermal gradient instead of a sharp discontinuity. The temperature profile insets of Figures 6 and 7 can be considered typical. Furthermore, only two salinity profiles were assumed, and these were derived from averaging only a few days' titration values for water samples taken once daily at discrete depths. Possibly an even more dominant factor is that the field of motion was not known. In view of these factors, it is perhaps surprising that the comparative values given in Table 1 agree as well as they do.

Another rough comparison of wave speeds predicted from (1) and (2) can be made using observed values. *Lee* [1961] reports 0.3 knot as the average speed of internal waves with periods between about 5 and 20 minutes and amplitudes up to 15 feet observed off San Diego, California, in a depth of 60 feet. Salinity gradients are normally small (personal communication of February 9, 1961). For purposes of comparison, the following values might be taken as representative of a two-layer system off San Diego: $T_1 = 17^\circ\text{C}$, $T_2 = 13^\circ\text{C}$, $h_1 = h_2 = 30$ feet, and $S_1 = S_2 = 33.5$ per mil. Substitution in (2) yields a wave speed of 0.65 knot. Since the average speed of 0.3 knot given by *Lee* [1961] includes no corrections for current velocity, and since the values for the computation were somewhat arbitrarily selected, it would seem possible that this calculated speed is in reasonable agreement with the measured average.

Several conclusions may be drawn from the general analysis as summarized graphically in Figures 8, 9, and 10. The passage of single

internal waves (which usually were in the form of a solitary wave trough) was common at TT4 during the two half-month observation periods when a marked negative thermal gradient persisted. These waves typically moved shoreward at speeds of about 1 knot (computed averages in each of the three principal sectors ranged from 0.95 to 1.00 knot). Deviations from this mean speed and direction were quite small, especially in view of the distorting effect of mean field motion (currents) for which no adjustments were made.

A somewhat quantitative estimate of the relationship between observed currents and wave signature speeds can be obtained from the data. From the averages previously given for a 2-week period not concurrent with internal wave observations, the range of average current speeds in each quadrant corresponds well to the variation in signature speeds. The standard deviations were 0.24 knot for the north-south components of current and 0.24 for the east-west components. The corresponding standard deviations for observed internal wave signatures were 0.38 knot (N-S) and 0.42 knot (E-W). If sampling errors, nonsimultaneous observations and other uncertainties discussed earlier are disregarded, these values, if statistically significant, seem to indicate that over half of the apparent wave speed variation is associated with the current regime. Of some importance is the fact that the current measurements were made 100 feet below MWL. It would probably not be realistic to attempt a more quantitative analysis using the data from Texas Tower 4; this must await the

completion of simultaneous observations of currents and internal waves over much longer periods of time.

Acknowledgments. The field work and most of the analysis for this study were completed while the author was employed at the U. S. Navy Hydrographic Office. Appreciation for release of information for publication is gratefully acknowledged. The work was completed at the Department of Oceanography and Meteorology of A. & M. College of Texas under Office of Naval Research contract N7onr2119(4). Personal appreciation is due Robert B. Elder, Howard V. Freilich, and W. A. Garth of the Hydrographic Office for encouragement and assistance.

REFERENCES

- Davis, P. A., and A. M. Patterson, The creation and propagation of internal waves, a literature survey, *Tech. Mem. 56-2*, Pacific Naval Laboratory, 1956.
- Defant, A., On the origin of internal tide waves in the open sea, *J. Marine Research, Sears Foundation*, 9(2), 111-119, 1950.
- Ekman, V. W., On dead water, *Scientific Results of the Norwegian North Polar Expedition, 1893-1896*, 6(1), 1-152, 1904.
- Fjeldstad, J. E., Interne Wellen, *Geofys. Publikasjoner, Norske Videnskaps-Akad. Oslo*, 10(1), 1-35, 1933.
- Haurwitz, B., Internal waves of tidal character, *Trans., Am. Geophys. Union*, 31(1), 47-52, 1950.
- Keulegan, G. H., Characteristics of internal solitary waves, *J. Research NBS*, 51, 133-140, 1947.
- Lee, O. S., Observations on internal waves in shallow water, *Limnol. and Oceanog.*, 6(3), 312-320, 1961.
- Long, R. R., Solitary waves in one and two-dimensional systems, *Tellus*, 8, 460-471, 1956.

(Manuscript received August 4, 1961.)

Periodicity of Earthquakes and the Invariance of the Gravitational Constant

W. J. MORGAN,¹ J. O. STONER,¹ AND R. H. DICKE

*Palmer Physical Laboratory
Princeton, New Jersey*

Abstract. Times of occurrence for a total of 1933 earthquakes are analyzed for periodicities. Results show no definite evidence for effects due to earth tides. Small indications of a date periodicity are assumed to be thermal in origin. A strong, statistically significant annual period is found, and the phase is substantially the same for northern and southern earthquakes. This periodicity is not wholly accounted for by temperature effects, wind-induced effects, and observer bias. The occurrence of this periodicity would be understandable if the gravitational constant were to vary as the earth-sun distance changes or as the earth's velocity relative to a preferred coordinate frame changes; however, the observed periodicity cannot be interpreted as conclusive support for such a hypothesis.

Introduction. It is usually assumed that the constants, including the gravitational constant, are invariants. On the other hand, arguments based upon the sizes of dimensionless constants [Dirac, 1938] and Mach's principle [Mach, 1917; Jordan, 1955; Dicke, 1959a, b] have been used to suggest that the locally measured value of the gravitational 'constant' may vary when expressed in atomic units. It has been suggested that the local value of the gravitational constant may be determined by a dimensionless quantity which is, in turn, determined by the distribution of matter in the universe. The local value of the gravitational constant on the earth should vary, as a result of changing distance to the sun, by roughly 10^{-10} , if these arguments are correct [Dicke, 1959a].

It is also possible that the local value of the gravitational constant is affected by the motion of the laboratory relative to distant matter [Dicke, 1959b]. Because of the presumed symmetrical distribution of distant matter, any such dependence would be an even function of the velocity relative to the moving coordinate system in which the universe appears isotropic. For small velocities the locally measured value of the gravitational constant would be expected to vary with velocity as

$$G_0[1 + \alpha(v/c)^2] \quad (v/c)^2 \ll 1. \quad (1)$$

¹ National Science Foundation predoctoral fellow.

Here it would be presumed that α should be either zero or a number with absolute value of the order of unity, say $0.1 < \alpha < 10$. Is there observational evidence which can be used to reduce α to such a small value that it can be assumed to be zero? The answer appears to be no.

First, it should be noted that the observed approximate isotropy of the distribution of galactic red shifts is an indication that the sun's velocity relative to this coordinate system is fairly small. (If it were not small, there would be many blue shifted galaxies in one direction.) The earth's velocity relative to the sun is also small. Hence the annual variation in $(v/c)^2$ is small, and there should be no gross effects but only very tiny effects associated with a velocity dependence of G .

Unfortunately, owing to the large sector of the sky obscured by galactic dust and the lack of observations for the southern hemisphere, the data for a velocity measurement based on galactic red shifts are too few. About all one can say is that, for those spatial directions for which data exist, the sun's velocity is zero plus or minus several hundred kilometers per second. To show the order of magnitude of the effects which need to be considered, it is assumed that the component of the sun's velocity lying in the plane of the earth's orbit is 100 km/sec and $\alpha = 0.2$. The resulting annual variation in G would have an amplitude of 1.3 parts in 10^9 .

It is difficult to rule out such small effects. The mechanics of the solar system would be

only slightly affected by a velocity dependence of the active gravitational masses of the individual planets. By *active gravitational mass* is meant the measure of the source strength of a body for the generation of a gravitational field (H. Bondi). The largest effect by far seems to be in the moon's motion. Here an annual variation of 2 parts in 10^8 in the active gravitational mass of the earth would lead to a periodic term in the moon's motion with an amplitude of 0.1 second of arc. Apparently there are observational difficulties with the moon which make it presently difficult to exclude an annual period with this small amplitude (Clemence, private communication). Other than the moon, the motion of Jupiter's satellites appears to be the next most hopeful source of astronomical data. The accuracy required is again almost hopelessly great.

The most promising laboratory apparatus for observing such an annual variation in G is the pendulum clock. The best pendulum clocks still running appear to be at the Paris Observatory. A recent examination of the data from these clocks for the years 1951-1957 by J. Peebles, a member of our group, has convinced us that an annual period of G with a fractional amplitude greater than 2×10^{-8} is not likely. There is some indication of a fractional amplitude of about 1×10^{-8} , but its statistical significance is questionable.

One fairly obvious geophysical effect which should accompany an annual variation in the active gravitational mass of the earth is an annual variation in its rate of rotation. This would result from the change in moment of inertia, in turn resulting from a periodic expansion and contraction due to a change in the gravitational interaction of the earth with itself. Using simplified earth models, Hess [1958] computed the resulting change in the rate of rotation. The result of his computation is that a change in G of 1 part in 10^8 would change the length of a day by about 1 part in 10^9 .

An annual variation of the length of a day with an amplitude of 3.6 parts in 10^8 is known [Markowitz, 1955]. This is usually considered to be due almost completely to a variation in the angular momentum of the earth's atmosphere [Mintz and Munk, 1954]. It is our opinion that there is not yet an accurate value for the effect of the earth's atmosphere. The

computations for the southern hemisphere based upon the Australia-New Zealand sea but the circulation pattern for the south hemisphere is highly irregular. It seems to that under these circumstances the computations for the southern hemisphere do not have reliability to allow other appreciable contributions to a variation in the earth's rotation to be excluded.

The present paper is concerned with another possible geophysical effect. An annual variation in G would result in the production in the earth of strains varying with an annual period. The superposition of this strain pattern with the very slowly changing strain patterns due to tectonic processes in fault zones might be expected to introduce an annual period in the rate of occurrence of earthquakes. To see that the order of magnitude is reasonable, it may be noted that an annual variation in G of 2 parts in 10^8 would result in earth strains of about 1×10^{-8} [Hess, 1958]. On the other hand, a reasonable estimate of the annual change in strain in the region of an active fault is 10^{-7} to 10^{-8} .

Unfortunately, geophysical processes are complicated and so poorly understood that it is difficult if not impossible to make an unambiguous prediction concerning the effect of a periodic stress on the rate of occurrence of earthquakes. Depending upon assumptions made, the rate of occurrence of earthquakes would have a periodic amplitude proportional to the $1, 0$, or -1 power of the frequency of the periodic stress.

With the assumptions that an earthquake occurs when some critical stress is exceeded and that only a linearly increasing stress of tectonic origin and some small periodic stress is involved it would be expected that the relative probability of earthquake occurrence would be proportional to the amplitude and the frequency of the periodic stress. However, if it is assumed that occasional short-period random disturbances occur, the relative probability of earthquake occurrence could be nearly independent of the frequency of the periodic stress. On the other hand, if rock creep plays a role in accumulating a stress across the fault plane, the amplitude of the periodic term in the rate of occurrence of earthquakes could vary over a limited frequency range inversely as the frequency of the periodic stress.

ould be noted that the stress distribution ed with a changing gravitational con- ould be spherically symmetric and would ll substantially vertical faults in the same t. On the other hand, tidal distortions of th are predominantly P_2 [$P_2 = \frac{1}{2} (3 \cos^2 \theta - 1)$, θ measured from the sublunar point] and to open some faults and close others, ing upon location and orientation. In the the effects of such tidal distortions upon uake rates, averaged over all the fault on the surface of the earth, may be small. y authors have investigated the periodic- earthquakes [Davison, 1938; McMurry, Stetson, 1935; Allen, 1936; Perrine, 1949]. rious reasons, most of the results of these ys were not suitable for application to the t problem.

la used. The data used in this study were from the earthquake tables in *Seismicity Earth*, by Gutenberg and Richter [1954]. authors have tabulated several thousand w, intermediate, and deep earthquakes by (shallow means depth of focus between 0 0 km, intermediate means depth of 70 to n, and deep means depth greater than 300 They have listed time of occurrence, loca- epicenter, and magnitude for all events, depths for intermediate and deep earth- s. Dates of occurrence range from 1904 to with a distribution by year as shown in e 1. The large variation is probably largely fluctuation in the rate of occurrence of

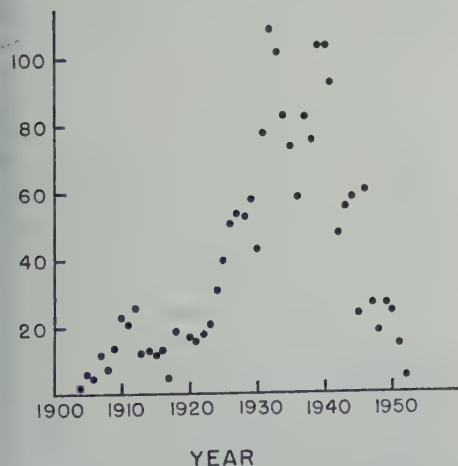


Fig. 1. Yearly distribution of tabulated earthquakes.

earthquakes, but it may be affected by human factors. Note the drop associated with both world wars.

All earthquakes of intermediate and deep focus listed in Table 18 of the reference were used in this analysis. Since few deep or intermediate earthquakes occur far out at sea, shallow earthquakes with epicenters far from land were used also. These latter included all occurrences listed in Table 17 under the regions: Atlantic Ocean, Pacific Basin, Southeastern Pacific, Eastern Pacific, Caribbean Sea, Indian-Antarctic Swell, and Indian Ocean. The authors felt that these events would provide a sample large enough to yield statistically significant results while suppressing the effects of local temperature-induced strains.

Method of analysis. $R(t)$, the rate of earthquake occurrence, is expressed as a sum of δ functions,

$$R(t) = \sum_{i=1}^N \delta(t - t_i)$$

where N is the number of earthquakes, t_i is the time of occurrence of the i th earthquake, $\delta(x) = 0$ for $x \neq 0$, and $\int_{-\infty}^{+\infty} \delta(x) dx = 1$. To examine $R(t)$ for a periodicity of period T , we first compress $R(t)$ so that the time varies only over the interval $0 \leq t < T$:

$$R(t, T) = \sum_{i=1}^N \delta[t - (t_i - n_i T)] \quad (2)$$

where n_i is an integer chosen to make $0 \leq (t_i - n_i T) < T$. Fourier analysis of this rate function yields

$$R(t, T) = b_0 + \sum_{k=1}^{\infty} a_k \sin \frac{2\pi k t}{T} + \sum_{k=1}^{\infty} b_k \cos \frac{2\pi k t}{T} \quad (3)$$

where the constants are given by

$$b_0 = \frac{N}{T}, \quad a_k = \frac{2}{T} \sum_{i=1}^N \sin \left(\frac{2\pi k (t_i - n_i T)}{T} \right), \quad (4)$$

$$b_k = \frac{2}{T} \sum_{i=1}^N \cos \left(\frac{2\pi k (t_i - n_i T)}{T} \right)$$

An IBM 650 computer was used to process the earthquake data. To speed up the calculations, each period T was subdivided into 100 equal parts, or categories. The time of occur-

rence of each earthquake was rounded off to the nearest one-hundredth of T , and the earthquake was placed in the appropriate category. Equation 2 then takes the form

$$R(t, T) = \sum_{i=0}^{99} N_i \delta(t - t_i)$$

where N_j is the number of events in the j th category and t_j is a characteristic time of occurrence of the j th category: $t_j = jT/100$. The constants of (4) which the machine computed were then

$$b_0 = \frac{N}{T}, \quad a_k = \frac{2}{T} \sum_{i=0}^{99} N_i \sin \frac{2\pi k j}{100}, \quad (5)$$

$$b_k = \frac{2}{T} \sum_{i=0}^{99} N_i \cos \frac{2\pi k j}{100}$$

We may write (3) in a more convenient form:

$$R(t, T) = \frac{N}{T} \left[1 + \sum_{k=1}^{\infty} A_k \cos \left(\frac{2\pi k t}{T} - \phi_k \right) \right] \quad (6)$$

where $A_k = (T/N)(a_k^2 + b_k^2)^{1/2}$ and $\phi_k = \tan^{-1}(a_k/b_k)$. The A_k 's and ϕ_k 's are the Fourier amplitudes and phases, respectively.

The longitude of each earthquake was combined with its time of occurrence (UT) to find its time of occurrence in local time. Local times of occurrence were used in the analysis.

If the results of Fourier analysis of a set of N independent random events are put into the form of (6), it may easily be shown that the distribution function $P(A_k)$, the probability per unit amplitude increment of getting a value A_k for any Fourier amplitude, is

$$P(A_k) dA_k = \frac{1}{2} N A_k \exp(-N A_k^2/4) dA_k \quad (7)$$

for N greater than about 8 [Rayleigh, 1899]. The mean-square amplitude resulting from any such analysis is

$$\overline{A_k^2} \equiv \sigma^2 = \int_0^{+\infty} A_k^2 P(A_k) dA_k = \frac{4}{N}$$

Again using the distribution function in an analysis of N random events, we find the probability of getting an amplitude for the k th harmonic greater than a previously assigned value A_{k0} to be

$$\begin{aligned} p(A_k > A_{k0}) &= \int_{A_{k0}}^{+\infty} P(A_k) dA_k = \exp(-N A_{k0}^2/4) \\ &= \exp(-A_{k0}^2/\sigma^2) \end{aligned}$$

Values of $p(A_k > A_{k0})$ are listed below for a few values of the quantity A_{k0}/σ .

A_{k0}/σ	$p(A_k > A_{k0})$
1	0.368
1.5	1×10^{-1}
2.1	1×10^{-2}
2.6	1×10^{-3}
3.0	1×10^{-4}

If in the analysis of earthquake occurrence the results a Fourier amplitude of $A_k > 2.6\sigma$ for the k th harmonic of interest, it is reasonably certain that the amplitude did not arise through random effects, and the associated periodicity may be associated with a physical, nonrandom cause.

Let us assume for the moment that the time of occurrence of earthquakes is correctly represented by the sum of three terms: a constant term, a periodic term having amplitude A_{kt} , a phase ϕ_{kt} , and a term representing random scatter. If a finite number of earthquakes is analyzed, differences between the measured amplitude A_k and the true amplitude A_{kt} , and between the measured phase ϕ_k and the true phase ϕ_{kt} will generally be found, due to random effects. The expected scatter of measured values can be found from the distribution function of (7). The rms deviations from A_{kt} and ϕ_{kt} for measurements on a large number of data samples for which $A_{kt} = 3\sigma$ are about 10 per cent and 10 degrees, respectively.

In previous publications by various authors, samples for analysis with N between 100 and 500 have generally been used. Also, regional subdivisions of the data have been made, sometimes reducing the number of events in a given analysis to less than a hundred. This has two obvious effects: (1) For equal subdivision of the original sample into n parts, there are now n places where statistical fluctuation could lead to a spurious large amplitude; (2) each category containing $1/n$ times as many events leads to spurious amplitudes $n^{1/2}$ times as large. So in another way, if the probability of obtaining an amplitude greater than a certain size through statistical fluctuation of a random sample

division of the sample into n equal sub-
samples will give for the probability that one
of the random subsamples has an am-
plitude greater than this same size

$$P' = 1 - (1 - P^{1/n})^n$$

Small P and large n , the effect on statistical
chance is drastic. For example, with 1933
earthquakes there is a probability $P = 5 \times 10^{-5}$
that a random sample will give an amplitude
greater than 0.143. However, divided on a re-
gion, depth, and/or strength basis into five
series each containing 386 earthquakes, the
probability P' that one category, or more, has
an amplitude this large or larger is $p' = 0.53$.
Because of the first effect, it is important that
in any study all results of investigations in
the neighborhood of the original data be published,
so that false conclusions may easily be drawn from
incompletely reported investigations. Because
of the second effect, it is important to include
all events as possible in searches for spec-
ific effects.

Results. Tables 1 and 2 show the results of
the general analysis. Five periods T were used,
and the first and second harmonics were com-
puted for each period. The five periods were:

Year	365.25 days
Solar day	1440.0000 minutes
Sidereal day	1490.4734 minutes
Synodic month	29.5305882 days
Anomalistic month	27.5545505 days

The synodic month is the time interval between
new moons; the anomalistic month is the inter-
val between successive perigees.

The division into subgroups was straightfor-
ward. Continental earthquakes are those inter-
mediate earthquakes having foci located far
from coasts (regions 25, 26, 27, 29, 30, 31, 47,
and 51 in Table 18). Shallow earthquakes all
have foci under water far from land, as was
shown earlier.

The zeros for phases are referred to the be-
ginning of the year for the yearly period and to
midnight for the solar day. (The sun is over-
head at the site of the earthquake for zero-hour
time.) Since only the yearly and daily periodicities
came out strongly, it was not thought nec-
essary to relate the phases of the three latter
periods to characteristic points in their respec-
tive cycles; however, December 31, 1200 UT,

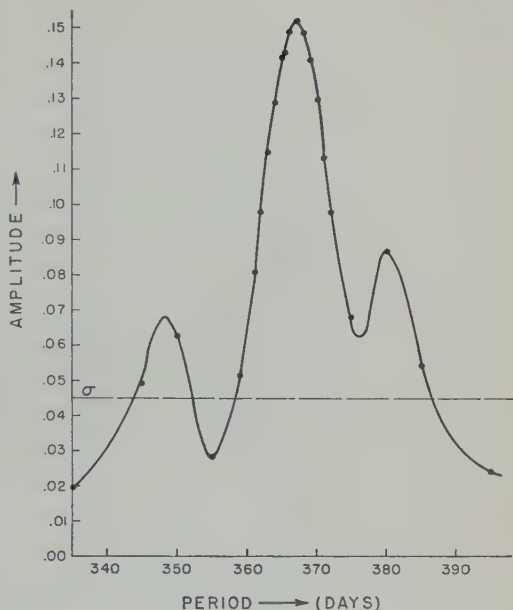


Fig. 2. Fourier spectrum of earthquake occurrence
in vicinity of yearly period.

1899, at 0° longitude, corresponds to zero time.
At this time the moon was 28.5 days past new.

The value of $p(A_k > A_{k0})$, calculated from
(8) for each amplitude, gives the probability of
obtaining that amplitude from analysis of a
random distribution of N events. These numbers
are consistently small only for the yearly period.

After completion of the survey described
above, a more intensive study was made of the
periods of most interest. The Fourier ampli-
tudes for the first and second harmonics were
computed for many periods T in the neighbor-
hood of the period of interest. All 1933 earth-
quakes were used each time, except in Figures
3 and 4.

Figure 2 shows a spectrum analysis for the
yearly period. Because of the regularity and
distinctive nature of the peak, there can be
no reasonable doubt as to the presence of a
yearly period in the occurrence of earthquakes.
The finite width of the peak is due to the fact
that the events analyzed are spread over a
finite time (see Fig. 1). The measured width
of the peak and the location of the side peaks
correspond to the earthquakes being concen-
trated in an approximately 30-year interval.

In the survey, the amplitude with the least
probability of being formed by chance occurred

TABLE 1. First Harmonic Amplitudes and Phases, with Associated Probabilities $p(A_k > A_{k0}) = \exp(-A_k^2/\sigma^2)$

	<i>N</i>	σ	Year			Solar Day			Lunar Day			Synodic Month			Anomalistic Month		
			<i>A</i> ₁	ϕ_1	<i>p</i>	<i>A</i> ₁	ϕ_1	<i>p</i>	<i>A</i> ₁	ϕ_1	<i>p</i>	<i>A</i> ₁	ϕ_1	<i>p</i>	<i>A</i> ₁	ϕ_1	<i>p</i>
All	1933	.046	.143	179°	.000048	.086	172°	.027	.029	211°	.67	.041	277°	.45	.035	124°	.54
All north	1026	.063	.199	165°	.000037	.091	184°	.12	.012	99°	.97	.013	319°	.96	.026	74°	.84
All south	907	.067	.101	210°	.0099	.086	157°	.18	.067	222°	.36	.077	270°	.26	.061	145°	.43
0 to 60 km																	
All shallow	550	.085	.129	164°	.10	.043	40°	.77	.022	172°	.94	.195	241°	.0053	.021	111°	.94
N. shallow	281	.120	.189	174°	.082	.011	222°	.99	.030	311°	.94	.251	255°	.012	.079	81°	.65
S. shallow	269	.122	.077	138°	.67	.099	41°	.51	.071	155°	.71	.159	217°	.18	.050	236°	.85
70 to 300 km																	
All inter.	1051	.062	.144	203°	.0045	.154	186°	.0019	.049	253°	.53	.066	23°	.32	.034	112°	.74
N. inter.	581	.083	.182	172°	.0081	.122	193°	.12	.015	174°	.97	.090	64°	.31	.044	140°	.76
S. inter.	468	.093	.173	245°	.030	.197	181°	.011	.109	262°	.25	.097	334°	.33	.038	70°	.85
>300 km																	
All deep	332	.110	.249	148°	.0059	.102	131°	.43	.084	140°	.55	.036	263°	.90	.073	147°	.64
N. deep	162	.157	.317	142°	.017	.150	153°	.40	.107	87°	.63	.081	65°	.77	.164	332°	.34
S. deep	170	.154	.188	156°	.22	.085	91°	.74	.126	175°	.51	.144	254°	.41	.299	150°	.023
All deep and inter.	1383	.054	.151	184°	.00036	.132	177°	.0023	.034	221°	.67	.046	14°	.48	.041	126°	.55
All continental	185	.147	.214	131°	.12	.097	216°	.65	.065	345°	.82	.165	193°	.28	.173	150°	.25

TABLE 2. Second Harmonic Amplitudes and Phases, with Associated Probabilities $p(A_k > A_{k0}) = \exp(-A_k^2/\sigma^2)$

	N	σ	Year			Solar Day			Lunar Day			Synodic Month			Anomalistic Month		
			A_2	ϕ_2	p	A_2	ϕ_2	p	A_2	ϕ_2	p	A_2	ϕ_2	p	A_2	ϕ_2	p
All	1933	.046	.069	347°	.10	.011	318°	.94	.028	262°	.68	.033	74°	.60	.037	291°	.52
All north	1026	.063	.110	4°	.044	.028	252°	.82	.035	208°	.73	.066	70°	.32	.061	274°	.39
All south	907	.067	.046	296°	.61	.031	27°	.80	.049	302°	.58	.008	206°	.99	.023	348°	.88
0 to 60 km																	
All shallow	550	.085	.101	71°	.25	.054	258°	.67	.045	215°	.76	.060	255°	.61	.019	121°	.95
N. shallow	281	.120	.230	50°	.024	.115	241°	.40	.054	15°	.82	.048	317°	.85	.072	3°	.69
S. shallow	269	.122	.090	173°	.58	.036	355°	.92	.145	207°	.24	.109	231°	.45	.100	163°	.51
70 to 300 km																	
All inter.	1051	.062	.076	317°	.22	.015	291°	.94	.011	190°	.97	.089	67°	.12	.062	281°	.36
N. inter.	583	.083	.132	338°	.077	.035	242°	.83	.109	182°	.18	.138	77°	.063	.107	243°	.19
S. inter.	468	.093	.063	243°	.63	.034	12°	.88	.112	0°	.23	.044	24°	.80	.091	348°	.38
>300 km																	
All deep	332	.110	.180	333°	.069	.102	61°	.42	.135	302°	.22	.040	150°	.88	.063	328°	.72
N. deep	162	.157	.082	259°	.76	.155	50°	.38	.156	313°	.37	.042	227°	.93	.094	328°	.70
S. deep	170	.154	.336	347°	.0082	.061	89°	.85	.122	288°	.53	.079	120°	.77	.034	327°	.95
All deep and inter.	1383	.054	.100	324°	.032	.019	33°	.88	.030	287°	.73	.070	75°	.19	.059	292°	.30
All continental	185	.147	.089	290°	.69	.117	199°	.53	.165	192°	.29	.108	128°	.58	.070	91°	.80

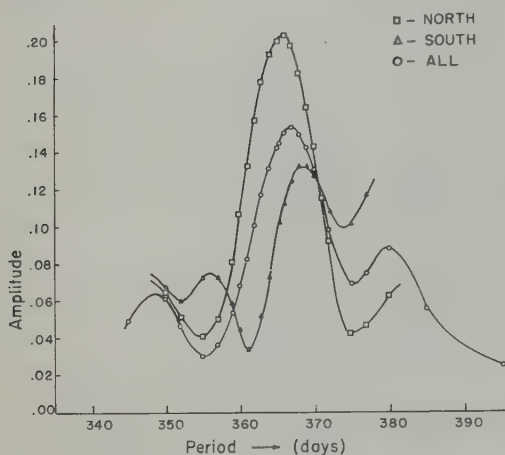


Fig. 3. Fourier spectrum of earthquakes in vicinity of yearly period.

for all northern earthquakes and the yearly period. A spectrum analysis was made of this subgroup. As expected, it exhibits the most symmetric peak because the effects of noise are relatively less. Figure 3 shows an analysis of all northern (1026 events) and all southern (907 events) earthquakes in the vicinity of the yearly period. The results, shown in Figure 2, are repeated for comparison.

The computer was programmed to calculate second-harmonic results simultaneously with the first-harmonic results. Figure 4 shows a spectrum analysis of the second harmonic of the yearly period for all northern, all southern, and all earthquakes.

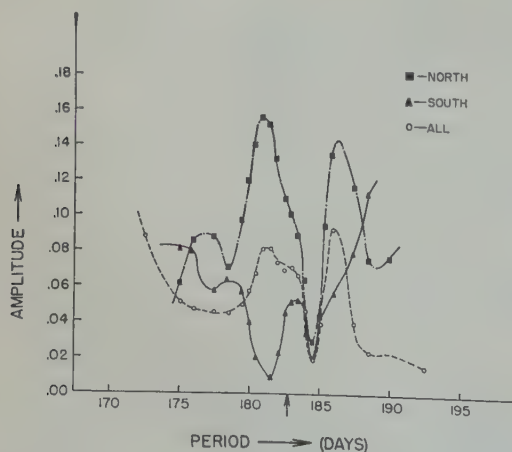


Fig. 4. Fourier spectrum of earthquakes in vicinity of half-year period.

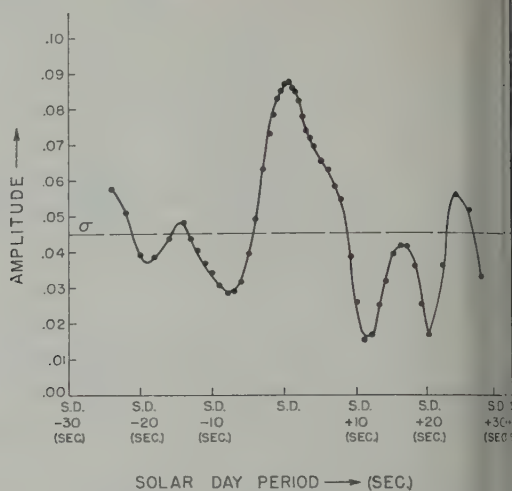


Fig. 5. Fourier spectrum of earthquakes in vicinity of solar-day period.

Figure 5 shows the spectrum for the solar day period. Again, the regular spacing for the side peaks corresponds to the earthquakes being concentrated in about a 30-year interval. Figure 6 shows the second harmonic of this, or the half-solar-day spectrum. Notice the similarity of Figures 2 and 5 and Figures 4 and 6.

Because of the strong tidal effects due to the moon, a half-lunar-day periodicity was expected even though the general survey gave null results. Spectrum analyses for the lunar and the half lunar day were made and are

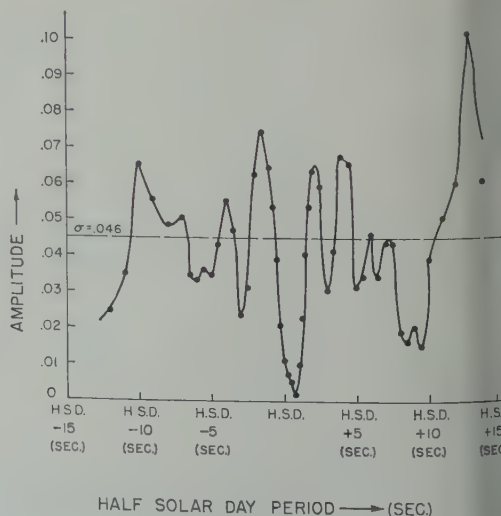


Fig. 6. Fourier spectrum of earthquakes in vicinity of half-solar-day period.

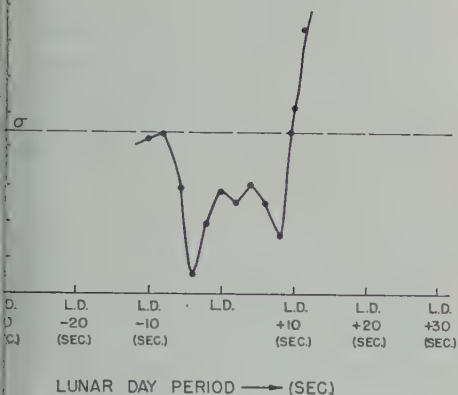


Fig. 7. Fourier spectrum of earthquakes in vicinity of lunar-day period.

in Figures 7 and 8. If a half-lunar-day were present, it would have a width of 2.5 seconds. No spectrum of the monthly was made because the survey showed the monthly amplitudes were small.

Discussion. The clear indication of a yearly, the less strong solar-day period, and very slight evidence for a lunar period is explanation. A first reaction is that the two periods have nothing to do with earthquakes, but merely mirror the periodicities in lives of investigators. The yearly effects of season, temperature, snow, and monsoons on habits could creep into the reduction of the possibility of a daily variation arising from coffee breaks, etc., seems remote.

With the hypothesis of an observer-induced periodicity in mind, the earthquake data were reduced as a whole for a 7-day period. The results were: amplitude = 0.039, $p(A_e > A_{10}) = 0.03$. Since a strong weekly period is not seen, they believe that the apparent small daily variations in earthquake occurrence are not due to changes in seismographers' efficiencies.

It is unlikely that the annual period is wholly due to this sort of cause, although it may be a part. There are rather substantial variations in amplitudes of the yearly fluctuations for different classes of earthquakes (from 0.1 to 1.0) and it is difficult to see why an observer-

induced periodicity should be different for different classes of earthquakes.

The fact that both periodicities involve the sun suggests that these periodicities are, at least in part, thermal in origin. The class of shallow earthquakes (under water, far from land) was included in order to have a check on thermal effects. There is no appreciable yearly or daily change in the bottom temperature of the mid-oceans, so thermal effects should not appear strongly in the occurrence of these earthquakes. Here indeed there is no significant amplitude of the solar-day period; however the yearly period has an amplitude and phase not significantly different from the class of all earthquakes. If the yearly period were primarily thermal in origin, the rates of occurrence in the northern and southern hemispheres should have opposite phases. It should be noted that the amplitudes of the yearly period are generally less in the southern hemisphere than in the northern, which suggests a superposition of a climatic effect opposite in phase for the two hemispheres and a larger global disturbance having the same phase everywhere. It is consequently possible that a significant part of the annual period is thermal in origin. In this connection, it is surprising that the amplitude of the annual period for deep-focus earthquakes is larger than the average. These earthquakes occur at depths

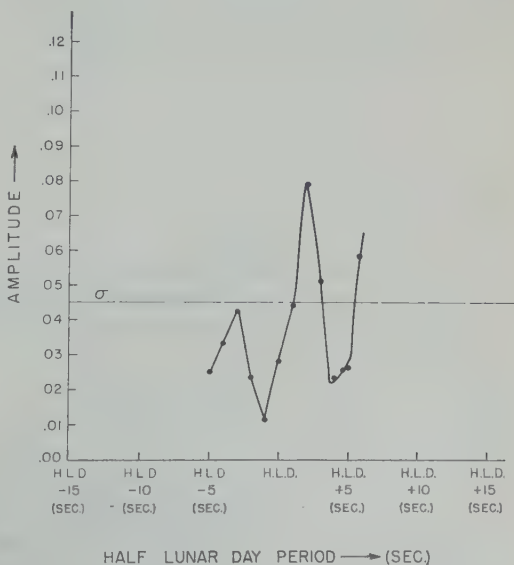


Fig. 8. Fourier spectrum of earthquakes in vicinity of half-lunar-day period.

greater than 300 km, and surface-induced strains would be expected to be small at such great depths.

Since the stresses of tidal origin due to the changing earth-moon distance are considerably greater than those due to the changing solar distance, one would expect that amplitudes of the periodicities induced in earthquake occurrence would be greater for the anomalistic month periodicity. This would seem to exclude the tidal effect due to a changing earth-sun distance as the cause of the observed yearly periodicity.

Periodic fluctuations in the heights of sea levels at coasts, resulting in periodic loading of the continental shelves, have been discussed by *Leypoldt* [1941] as possible causes of earthquake periodicities. These fluctuations are irregular and have amplitudes of about half a foot. In general, water levels are highest in the summers of the respective hemispheres. *Leypoldt* has attributed this circumstance to meteorological causes which are out of phase in the northern and southern hemispheres, for instance, the storing and release of snow, ice, and ground water. A large part of the sea level fluctuations may be due to thermal expansion of the oceans in the summer; this would cause a rise in sea level without an associated stress increase, since only the density and not the mass of water over continental shelves would be changed. This phenomenon may require further study; it does not, however, explain the fact that the rates

of occurrence have nearly the same phase in northern and southern hemispheres, nor the periodicity in mid-ocean earthquakes.

The present statistical results are not incompatible with those of *Landsberg* [1940]. *Landsberg* also proposed that the movement of a masses between northern and southern hemispheres causes a change in the moment of inertia of the earth, with consequent wandering of the axis of rotation. The observed periodicity of earthquakes is attributed to the stresses arising from these changes. However, the stresses in the earth which would be involved in such movement of the axis have been estimated by the authors (see Table 3) to be several orders of magnitude less than the lunar tidal stresses as calculated by *McMurry* [1941].

Another possible cause of an annual periodicity is the stress pattern associated with the annual change in atmospheric angular momentum in the northern and southern hemispheres. These changes are opposite and require a twist along the earth's axis for their production. Here also the resulting shear appears to be too small to be important.

As a check on other possible correlations of earthquake occurrence with astronomical phenomena, the data were examined for a sidereal day period. The resulting amplitude was 0.03 for 1933 earthquakes ($\sigma = 0.046$).

Studies have been made to determine the stresses in a solid sphere resulting from the tidal attraction of another sphere [*Hoskins*, 1920].

TABLE 3. Estimated Effects of Various Stress-Producing Mechanisms

Source of Stress	Assumptions Made in Calculation	Stress, dyne/cm ²
Lunar tidal stress	Compressive stress at sublunar point, depth 200 km, μ = modulus of rigidity = 1.15×10^{12} dyne/cm ² [<i>McMurry</i> , 1941, p. 75].	3.4×10^4
Tidal change in sea level over continental shelf	1-meter tide, stress computed at depth of 200 km under coast line [<i>McMurry</i> , 1941, p. 54].	6×10^4
Atmospheric pressure change	1-inch barometric change (1-inch barometer change is equivalent to 30-cm ocean tide).	2×10^4
Change in gravitational constant	$\delta G/G = 10^{-8}$ [<i>Hess</i> , 1958].	3.4×10^4
Microseism	Compressional wave, 10-sec period, 2-micron amplitude, 3000-meter/sec velocity, Young's modulus = $2.5 \times$ modulus of rigidity = 2.87×10^{12} dyne/cm ² .	1.2×10^4
Movement of axis of rotation of earth	Point at which axis cuts surface of earth moves uniformly 10 cm in 1 year. Cause is uniform stress over $\frac{1}{2}$ of earth's cross section through equator.	0.3

the results have been applied to the earth-system [McMurry, 1941]. McMurry, assuming a homogeneous earth, gives for the tidal stresses at a depth of 200 km:

$$rr = \mu c(-.000 - .008 \cos 2\theta)$$

$$\theta\theta = \mu c(+.044 - .221 \cos 2\theta)$$

$$\phi\phi = \mu c(-.132 - .044 \cos 2\theta)$$

$$r\theta = \mu c(-.026 \sin 2\theta)$$

$$\theta\phi = r\phi = 0$$

Let $\mu c = 1.93 \times 10^5$ dyne/cm² and θ is measured from the sublunar point (see Fig. 9). At a depth of 0 km, the rr and $r\theta$ components of the stress tensor equal zero while the $\theta\theta$ and $\phi\phi$ components are approximately the same as shown above. To simplify the discussion which follows, we shall suppose that only the largest stress is important in triggering earthquakes. We take $\theta\theta = \mu c(-0.2 \cos 2\theta)$; all other stresses equal zero.

Estimates of the magnitudes of forces due to the loading of continental shelves and atmospheric effects have been made by McMurry. Stresses arising from a $\delta G/G$ fluctuation of the earth's gravity have been estimated by the authors, as have stresses due to microseisms, which have been considered as possible earthquake triggering agents. A summary of these estimates is given in Table 3.

The amplitudes of periodic tidal strains are of the order of the estimated yearly increase in strain at a fault due to normal tectonic processes. This indicates that in any study of periodicity the lunar-day period should show up prominently. Other effects influencing earthquake

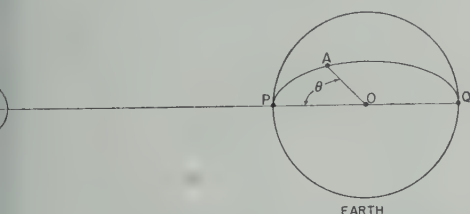


Fig. 9. The most prominent stress in the earth due to lunar attraction. Let P be the sublunar point, Q the antipodal point, O the earth's center, A a point of interest in the upper mantle or crust, PAQ a great circle. The stress at A is compression or tension directed along PAQ and of magnitude $S = 4 \times 10^4 \cos 2\theta$ dyne/cm².

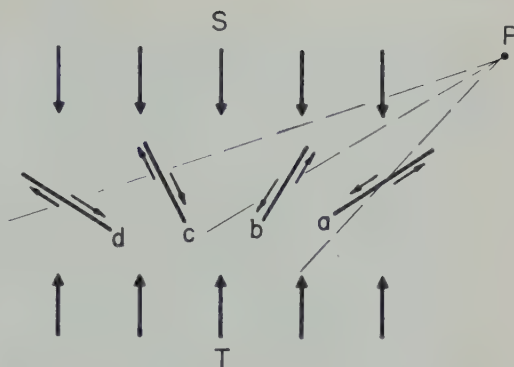


Fig. 10. Illustration of opposite tendencies of lunar stresses toward triggering earthquakes. Let P be the sublunar point; then there is a compressive tidal stress at each point directed towards P . Let tectonic processes be compressing the crust between S and T . Several possible fault planes are shown, and the arrows indicate the direction the S - T stress would cause these to slip. The additional stresses caused by the moon would tend to aid slip at a and to hinder slip at b , c , and d .

occurrence, such as eccentricity of the earth's orbit, a velocity dependent gravitational constant, or increased microseism activity, should be small in comparison with tidal effects if it is assumed that an earthquake occurs when a critical stress is exceeded at a fault. The fact that a strong lunar-day periodicity was not found implies that the triggering by tidal stresses is independent of the hour angle. We propose the following mechanism to account for this independence.

Figure 10 shows several possible fault orientations with respect to the sublunar point P and the earthquake-causing stress having a tectonic origin. If there were strike slip for each of these faults, they would slip as shown by the arrows when the stress (compressive in this example) due to tectonic origin exceeded the respective critical stress at each fault surface. The compressive tidal stress would tend to cause slip at fault a , to hinder slip at faults b and d , and to have little effect at fault c . These faults then, would be released preferentially at different positions (hour angles) of the moon. If fault planes and tectonic stresses are oriented at random, correlations of earthquake activity with either lunar or solar tidal forces might not be seen.

Since so little information on the orientations of fault planes and the directions of slip for

these planes is available, it is not possible to get a large body of data which would be homogeneous from this point of view. If one assumes that the proposed cancellation does occur, then it is reasonable to expect correlations between earthquake occurrences and small-amplitude stress fluctuations such as would be caused by a varying gravitational constant.

Benioff [1951] has assigned certain elastic and plastic characteristics to rocks on the basis of his study of earthquake sequences. His results suggest to the authors that one cannot speak of a definite critical stress level at a fault, which when exceeded immediately precipitates an earthquake. Rather, there may be a stress level below which all rock strains are elastic, considered over times of the order of a hundred years (an earthquake building time), and above which plastic flow begins. If the stress is held constant above the critical level, the rate of plastic flow may be nearly constant at first but will increase with time until rupture occurs and a macroscopic fault slippage takes place. Although no extensive consideration of this point has been made, it seems likely that the sensitivity of earthquake occurrence to externally caused periodic stresses could be strongly frequency dependent, favoring the correlation of occurrences with either short- or long-period stresses according to the mechanical properties of rock near the critical stress.

Summary and conclusion. As observed on the earth, an annual variation in the gravitational constant of the order of one part in 10^{10} would accompany a change in solar distance for certain cosmologies [*Sciama*, 1953; *Jordan*, 1955; *Brans and Dicke*, 1961]. An annual variation of one part in 10^8 due to a motional effect is also possible [*Dicke*, 1959b]. Such effects, if they occurred, would lead to purely radial stresses in the earth. The annual period in such stresses could lead to an annual period in the rate of occurrence of earthquakes. Because of the purely radial character of these stresses, a stronger correlation with earthquake frequency would be expected than for the P_2 type of tidal stress. Furthermore, for an annual period, the same phase would be expected for the northern and southern hemispheres.

A strong annual period is found in earthquake frequency, and the phase of this period is substantially the same for both northern and

southern earthquakes (earthquake occurrence rate is a maximum in June and July). Surprisingly, the amplitude of the period is greatest for deep-focus earthquakes, which makes a surface effect such as temperature seem unlikely as a causal explanation. The fact that the phases in the north and south are the same also seems to make a seasonal temperature variation unlikely as an explanation. This conclusion is also supported by the presence of the period for mid-ocean earthquakes. The absence of a half-year period makes unlikely a seasonal temperature effect on a global scale. A yearly period due to wind-induced stresses is a possibility, but these computed strains are very small.

It was suggested to us by F. Press that the fact that most seismic bulletins are published annually could lead to a fictitious annual period. This is certainly true, and it is difficult to exclude such a possibility. The best statistical tests, it seems to us, would need to be made individually by the seismographers who publish data in this way. In lieu of such tests it is possible to do little more than note that probability bias, proportional to the age of the data at the time of publication, would lead to a saw-tooth probability bias curve with an annual period. For such a probability distribution the amplitude of the semiannual period should be half that of the annual period, and its phase angle should be $\phi_2 = 2\phi_1 + 90^\circ$. There are indications of a statistically significant amplitude for the second harmonic of the annual period, but its phase is badly off.

Professor Press also suggested that, because seismographic stations are mainly in the north, heightened microseismic activity during winter months could bias data by causing weak earthquakes to become lost in the seismic noise. This could certainly be an important effect. If so, it might be expected that the amplitude of the annual period would be correlated with the strength of the earthquake. *Dasgupta* [1933] found such a correlation, but it was small and had the wrong sign. Strong earthquakes had the larger annual amplitude. The three classes of earthquakes we examined had very nearly the same average strength (6.34, 6.39, 6.40) but they differed considerably in the amplitude of the annual period. This possible explanation of the annual period seems not to be supported by the statistics.

all indications of a solar-day periodicity assumed to be thermal in origin.

The only evidence for an effect associated with earth tides is the suggestion of the synodic month period, with a small but possibly significant amplitude, for the class of shallow earthquakes. The statistical significance of this is questionable, for there are many cases of earthquakes in which a chance period could occur. Also, the absence of the lunar-day period makes its interpretation difficult.

It is concluded that the annual period exists as a statistically significant period. There is not yet a convincing conventional explanation for this period. An annual variation in active gravitational mass of the earth could cause such a period, but the presence of the period cannot be interpreted as strong support for the hypothesis. Conversely, if the postulated annual variation in the gravitational interaction should be inhibited in some other way, this would be a very reasonable causal mechanism for the annual period in earthquake frequency.

Acknowledgments. We wish to thank the personnel of the Statistical Techniques Research Laboratory, especially Pamela Johnson and Dr. Roger Ham, for their assistance with the computational problems and suggestions offered during the framing of the problem. Thanks are also due to Dr. Turner, who did some of the final calculations, and to Dr. Peebles for his analysis of the performance of the Paris pendulum clocks. Finally we wish to thank Professor Frank Press for reading the manuscript and making several helpful suggestions. This research was supported in part by research grants from the Office of Naval Research and the Atomic Energy Commission.

REFERENCES

- Benioff, H., Earthquakes and rock creep, I, Creep characteristics of rocks and the origin of after-shocks, *Bull. Seism. Soc. Am.*, **41**, 31, 1951.
- Davison, C., *Studies on the Periodicity of Earthquakes*, Thomas Murby and Company, London, 1938.
- Dicke, R. H., Gravitation—an enigma, *American Scientist*, **47**, 25, 1959.
- Dicke, R. H., New research on old gravitation, *Science*, **129**, 621, 1959.
- Dirac, P. A. M., A new basis for cosmology, *Proc. Roy. Soc. London A*, **165**, 199, 1938.
- Gutenberg, B., and C. F. Richter, *Seismicity of the Earth*, 2nd ed., Princeton University Press, Princeton, 1954.
- Hess, G., The annual variation of the length of the day as evidence relating to a theory of gravity, Physics senior thesis (unpublished), Princeton University, 1958.
- Hoskins, L. M., The strain of a gravitating sphere of variable density and elasticity, *Trans. Am. Math. Soc.*, **21**, 1, 1920.
- Jordan, P., *Schwerkraft und Weltall*, 2nd ed., Vieweg, Braunschweig, 1955.
- Landsberg, H., Seasonal pressure-changes and earthquake-occurrence, *Trans. Am. Geophys. Union*, part 2, 227-228, 1940.
- Leypoldt, H., Sea-level changes as trigger forces, *Bull. Seism. Soc. Am.*, **31**, 233, 1941.
- Markowitz, W., The annual variation in the rotation of the earth, 1951-54, *Astron. J.*, **60**, 171, 1955.
- McMurtry, H., Periodicity of deep-focus earthquakes, *Bull. Seism. Soc. Am.*, **31**, 33, 1941.
- Mintz, Y., and W. Munk, The effect of winds and bodily tides on the annual variation in the length of day, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl.* **6**, 566, 1954.
- Perrine, C. D., Periods in earthquake activity, *Bull. Seism. Soc. Am.*, **39**, 109, 1949.
- Rayleigh, John William Strutt, 3d baron, *Scientific Papers*, vol. 1, Cambridge University Press, Cambridge, pp. 491 ff., 1899.
- Sciama, D. W., On the origin of inertia, *Monthly Notices Roy. Astron. Soc.*, **113**, 34, 1953.
- Stetson, H. T., The correlation of deep-focus earthquakes with lunar hour angle and declination, *Science*, **82**, 523, 1935.
- W. W., The lunar triggering effect on earthquakes in southern California, *Bull. Seism. Soc. Am.*, **26**, 147, 1936.
- W. W., C., and R. H. Dicke, A scalar theory of gravitation and Mach's principle, to be published 1961.

(Manuscript received June 14, 1961; revised August 26, 1961.)

Some Characteristics of Surface Gravity Waves in the Sea Produced by Nuclear Explosions¹

WM. G. VAN DORN

*Scripps Institution of Oceanography
La Jolla, California*

Abstract. Low-frequency dispersive gravity waves produced by nuclear explosions at Bikini, Marshall Islands, were recorded at four distant island stations. The results of these wave measurements are compared with those predicted by linear theory, and good agreement is observed in the nature of the dispersion and the rate of amplitude decay with distance. The wave system associated with the large tsunami of March 9, 1957 is also considered. The dispersion for both types of disturbances was virtually identical, in agreement with the theoretical argument that the dispersion of a centered wave system is independent of the nature of the source disturbance. In analyzing the rate of amplitude decay, it was found necessary to correct the observations for enhancement due to scattering by the islands upon which the recording stations were located, showing that even relatively small islands are effective as scatterers. These experiments show rather conclusively that tide-gage records of tsunami-like disturbances are grossly misleading insofar as the characteristics of the wave systems in the open sea are concerned.

I. INTRODUCTION

This paper presents the results of measurements of low-frequency surface gravity waves in the sea produced by large nuclear explosions at Bikini atoll, Marshall Islands, during the spring and summer of 1956, as a part of the scientific field program carried out under operational Redwing. The propagation and dispersion relationships deduced from these data, as well as the relative rate of amplitude decay with distance, have a direct bearing on the validity of present hydrodynamic theory and its generalization to the problem of tsunami waves, which is a part of the same part of the amplitude-frequency spectrum of surface water waves.

The Redwing tests afforded a unique opportunity for the study of long waves from an impulsive disturbance on a scale which cannot be duplicated in the laboratory—or even in a more extensive field experiment—because of the unsolvable scaling problems. Such an experiment would require the generation of waves having wavelengths large, and amplitudes very small, compared with the water depth, but still capable of being resolved above normal background at distances large compared with the wavelength. Such nuclear tests under these conditions may

never be repeated, these data themselves are probably unique. Moreover, because the analysis presented here still leaves some uncertainty as to the most precise interpretation of the results, these data are tabulated herein in detail.

II. EXPERIMENTAL PROCEDURE

The wave observations reported here were obtained from four explosions on or within the confines of Bikini atoll, as listed in Table 1 together with the respective shot times and ground coordinates, as reported by *Griggs and Press* [1959], and as shown in Figure 1. Continuous long-period wave recordings throughout the test series were made at manned observation stations on Eniwetok atoll and on Wake and Johnston islands, and intermittent recordings of several days duration were also carried out at Ailinginae atoll from an autonomous station which was visited by ship before each test of interest. The geographical distribution of island stations and their great-circle distances from Bikini are shown in Figure 2, which also shows 'contours' of the distance traveled by a very long (\sqrt{gh}) wave in 20-minute increments originating at Bikini. These islands were selected from a rather limited list as being the best compromise among the many requirements for an optimum instrument base, such as accessibility, adequate communications, small size, topo-

¹Contribution from the Scripps Institution of Oceanography, New Series.

TABLE 1. Shot Times and Ground Coordinates for Nuclear Explosions at Bikini Atoll, Operation Redwing

Event	Date	UT	Latitude	Longitude
Cherokee	5/20/56	1750.6	11°40.1'N	165°23.7'E
Zuni	5/27/56	1756.0	11°29.8'N	165°22.1'E
Navaho	7/10/56	1756.0	11°39.8'N	165°23.2'E
Tewa	7/20/56	1746.0	11°40.5'N	165°23.2'E
Tsunami	3/9/57	1422	51.3°N	175.8°W

graphic isolation, and geographical distribution. At each observation island an instrument site was selected having a direct exposure toward Bikini and well removed from bays, inlets, and major topographic singularities which might result in anomalous amplitude or phase alterations of the incoming signals. At each site a long-period wave recorder was installed [Van Dorn, 1960] and connected to a plastic hose extending over the reef to a depth of about 30 feet. The amplitude response to a unit sine-wave input of variable frequency for these recorders (Fig. 3) was adjusted to be peaked at about 750 sec to give the greatest selective discrimination against swell and surfbeat. Figure 3 also shows the phase shift as a function of wave period.

Still greater discrimination against swell was possible at the option of local operators by interchanging capillaries in the hydraulic filter. Two optional amplifier sensitivities were also provided to accommodate the anticipated rapid decay of wave amplitude with range. Despite all these precautions, the wave signals were swamped at one or more stations during the early part of the test series by very high swell conditions, and the Eniwetok station habitually exhibited high surfbeat in the period range 3 to 5 minutes, so that eight out of twenty of the records were considered unreadable in the final analysis of results.

The local operators were instructed to place a calibration mark on the records at least once daily and to note the time of the calibration to the nearest minute, having first checked the watches against radio time. Such time and calibration control was not possible at Ailingi except at the beginning of the record. Since the recorder usually ran out of paper before the station could be revisited, the actual arrival times of the waves observed were uncertain by several minutes, although the times between the separate crest arrivals could be determined to much higher accuracy.

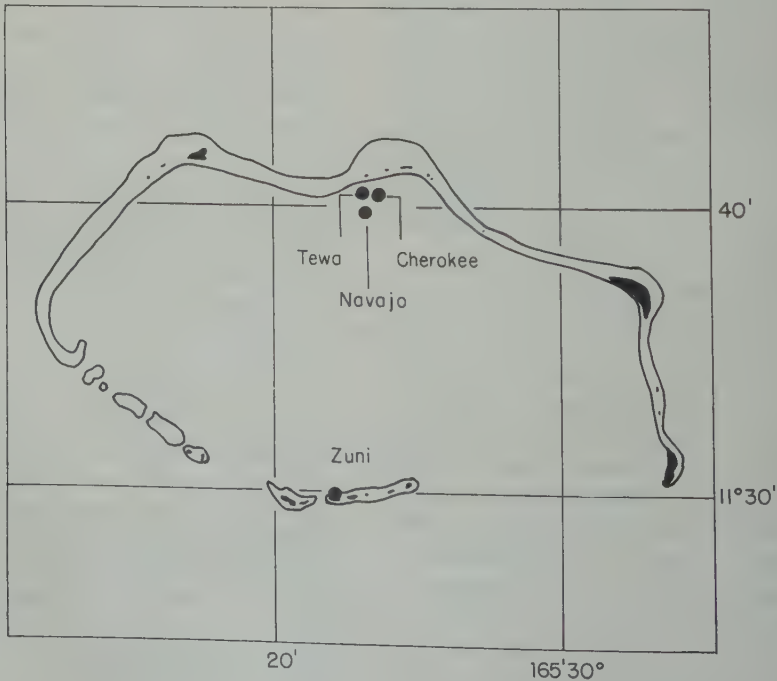


Fig. 1. Bikini atoll, showing shot coordinates for operation Redwing.

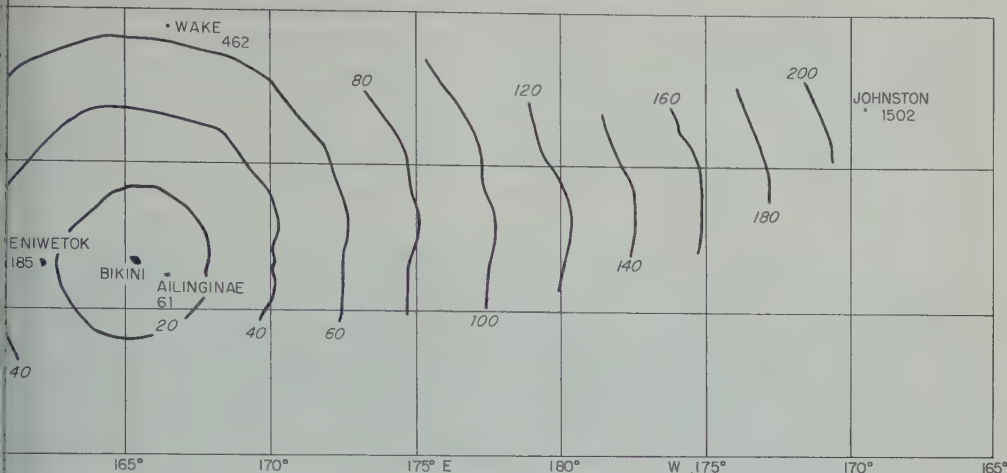


Fig. 2. Distribution of wave recording stations and their great-circle distances in nautical miles from Bikini atoll. Contours give instantaneous positions of wave front originating at Bikini at 20-minute intervals.

Following completion of the Redwing test, the Wake and Johnston stations were resumed in operation, and the Eniwetok and Ailinginae stations were removed to Canton Island and Kona Hawaii under a continuing research contract with the Office of Naval Research for the study of impulsively generated

waves. It was hoped to record waves from at least one large tsunami, as well as to attempt to determine the lower limit of earthquake energy responsible for tsunami production by examining the records for tsunami-like waves after each large earthquake. Despite frequent breakdowns—mostly due to the inability of

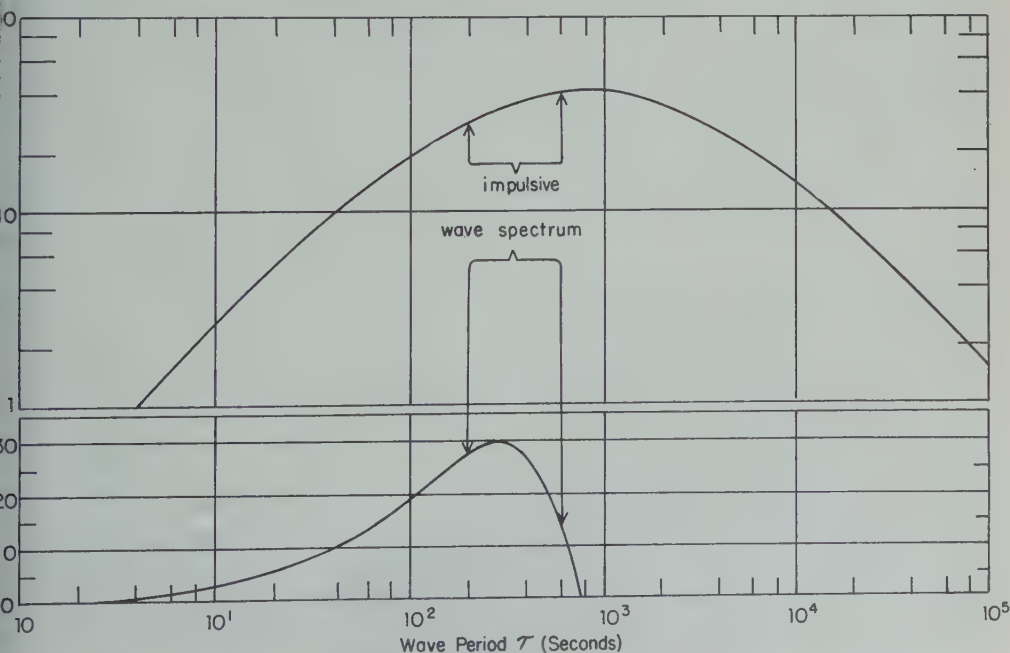


Fig. 3. Amplitude response and phase lag versus wave period for the long-wave recorders. All the waves observed fall within the indicated portion of the wave spectrum.

temporary stations to survive storm waves—these efforts were rewarded by obtaining a complete record of the large tsunami of March 9, 1957, at Wake Island, the other stations being out of service and in the process of remodeling at the time. It is fortuitous that Wake was one of the original Redwing stations, since a direct comparison of waves from two very different sources and modes of generation was pos-

sible. Pertinent data from this tsunami record are included in the following section, and the coordinates of the tsunami epicenter are listed in Table 1.

III. RESULTS OF WAVE MEASUREMENTS

Photographs of the original wave records are reproduced in Figures 4 through 8. Each record consists of normal background which is mod-

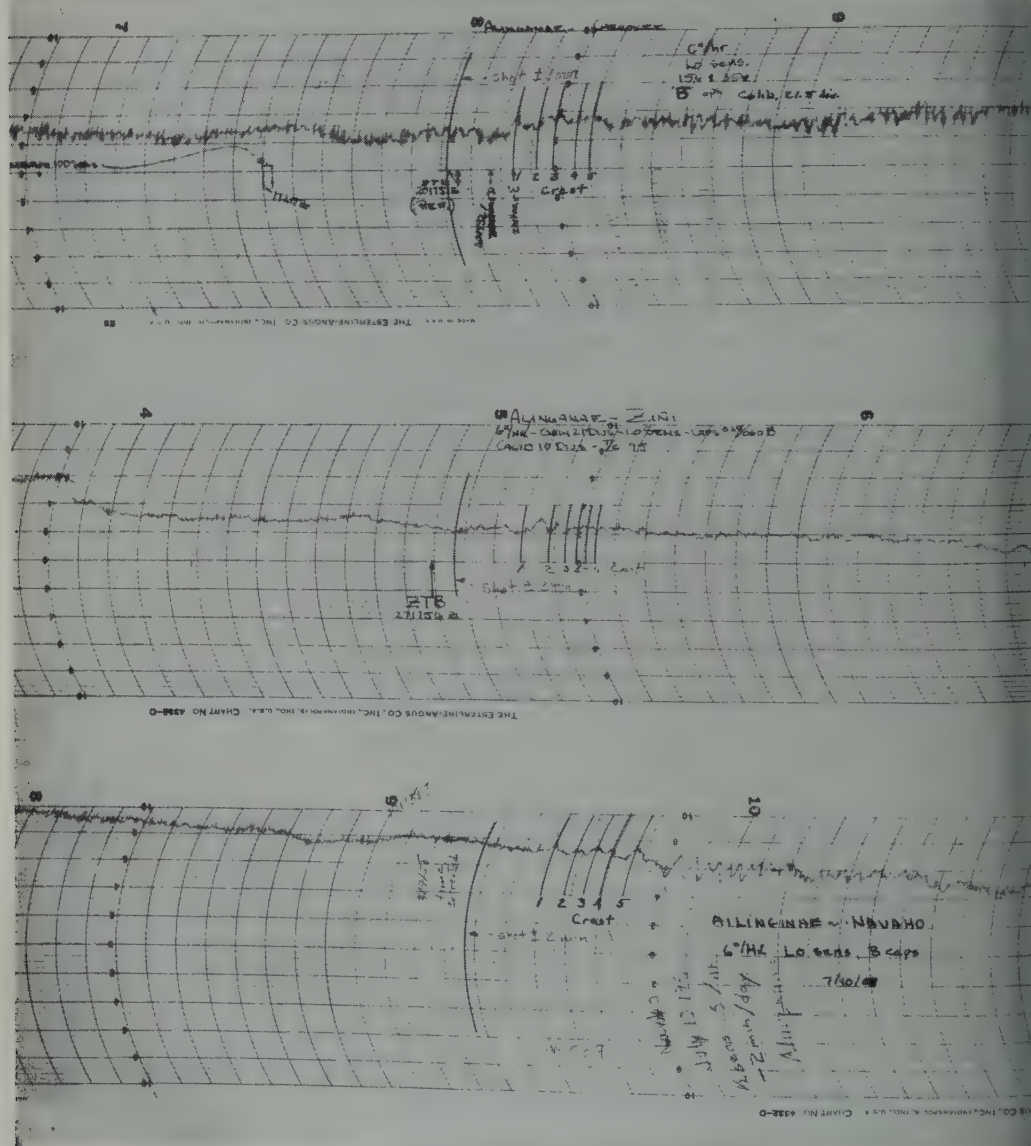


Fig. 4. Wave records from Cherokee, Zuni, and Navaho shots at Ailinginae atoll. Time scale is 5 min/division.

later by a dispersive wave train. The time is 5 minutes per division on all the bomb records and 10 minutes for the tsunami record. Clearly discernable, the first six individual arrivals are numbered in order of arrival. The arrivals from Ailinginae are somewhat irregular

and poorly resolved, partly because the instrument was set at low sensitivity and adjusted to give greater swell filtering, in expectation of large signals. The records for Navajo and Cherokee shots exhibit a secondary train of waves of larger amplitude than the first arrivals; the

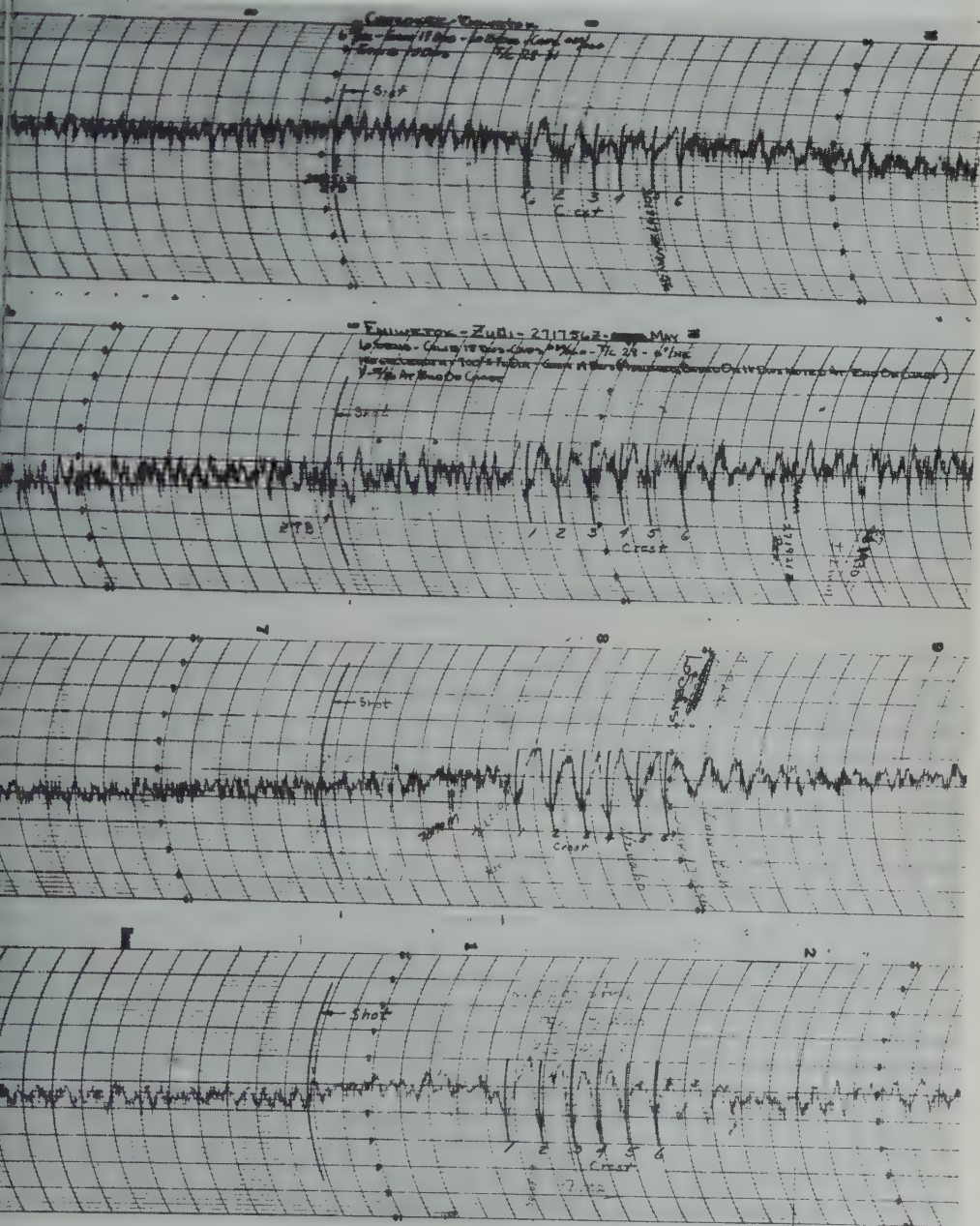


Fig. 5. Wave records from Cherokee, Zuni, Navaho, and Tewa shots at Eniwetok atoll. Time scale, 5 min/division.

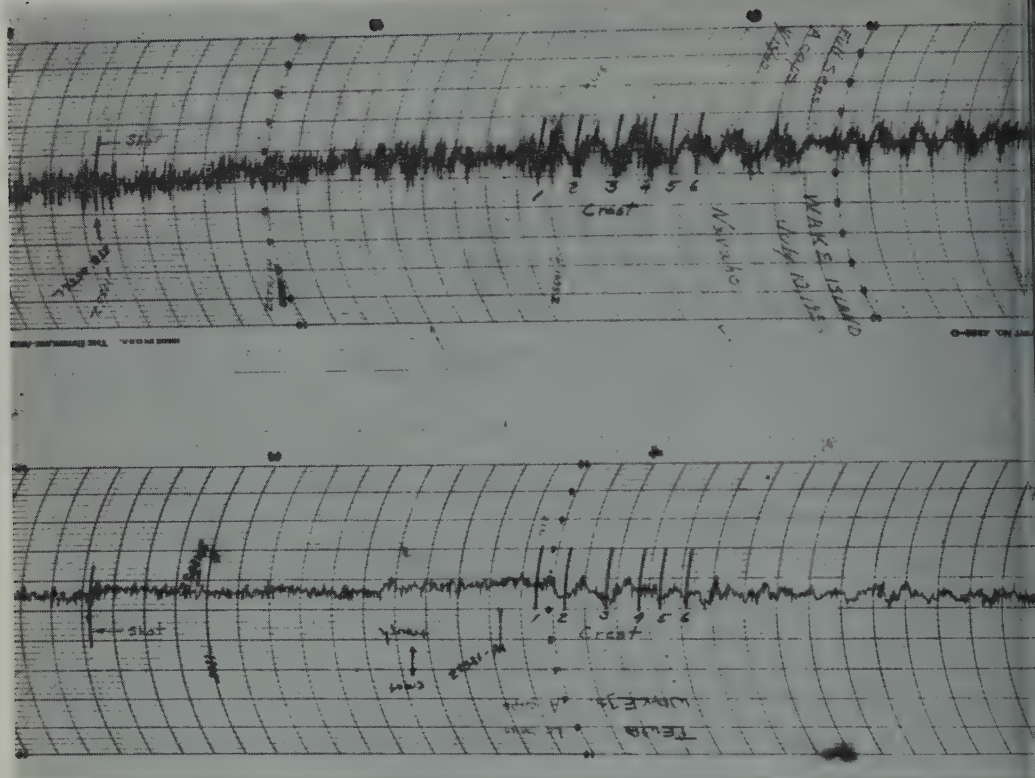


Fig. 6. Wave records from Navaho and Tewa shots at Wake Island. Time scale, 5 min/division.

latter are clearly waves generated within the lagoon leaking out of the wide southern pass in the direction of Ailinginae. There was good evidence that no lagoon waves overtopped the atoll rim in other directions and that the waves observed at the remaining stations were generated by that portion of the atmospheric impulse which extended beyond the atoll rim over the open sea. The wave trains observed at all stations beyond Ailinginae were remarkably coherent from test to test, so much so that at Eniwetok the records from the four tests can be overlaid and seen to correspond crest for crest with a time discrepancy of less than 1 minute. This result is the more remarkable when it is considered that Cherokee test took place at an elevation of 4300 feet above the atoll, whereas the remaining tests took place at sea level. Only the waves from Navajo and Tewa tests were clearly resolvable from the normal background at Wake and Johnston islands, but the same sys-

tematicity applied to the early waves from these tests at both stations.

Qualitatively, at all stations the first significant departure of the trace from normal background was a crest, and the highest crest occurred farther back in the train with increasing distance from Bikini. At Ailinginae the largest crest was the first, and at Johnston it was third in order of arrival. The wave trains were dispersive in that the initial wave had the longest period, and the period systematically decreased with time. After the passage of five or six waves, this regularity broke down into a more-or-less confused jumble of random oscillations. Therefore all analysis of these data was confined to the first six crest arrivals.

Wave arrival times. The arrival times (estimated to the nearest 0.1 min) of the first troughs and crests are listed in Table 2, the same ordering system being used as for Figures 7 through 8. Except for Ailinginae, the tr-

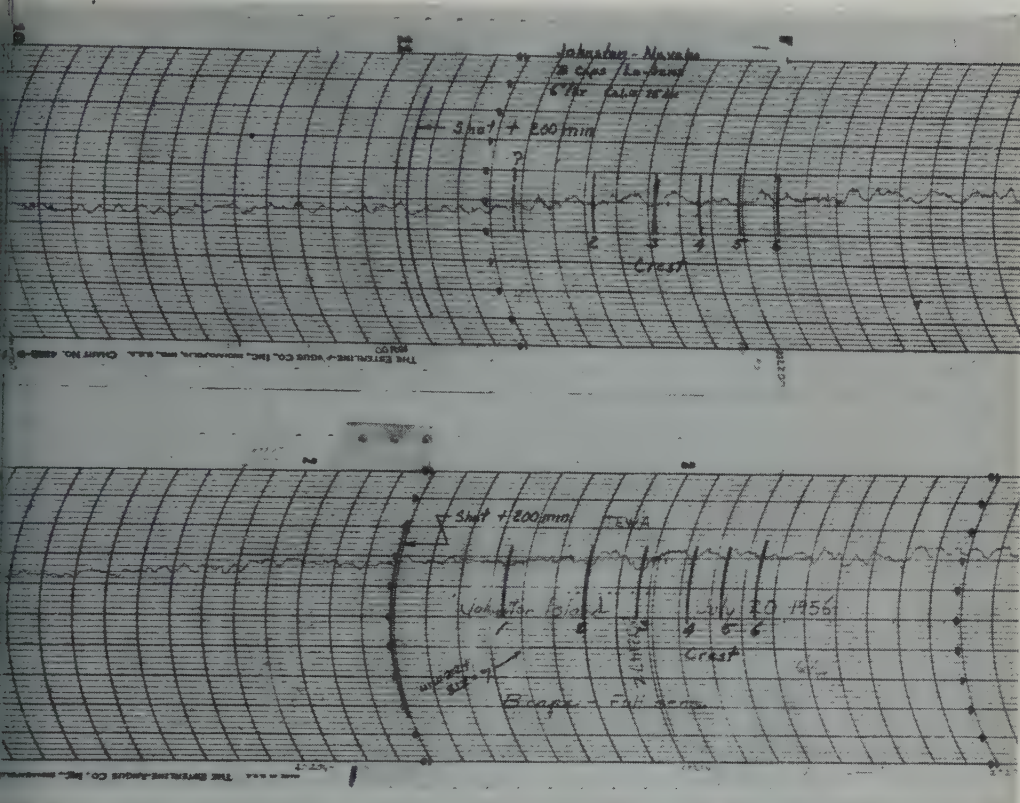


Fig. 7. Wave records from Navaho and Tewa shots at Johnston Island. Time scale, 5 min/division.

es to the various islands are large compared in the time error introduced by neglecting short distances between the points of origin the several tests. Because of the excellent erence between the various records at a single ion, the mean values of the travel times are

also shown for further computational use. The departure of individual arrivals from the mean is of the same order (1 min) as the over-all timing accuracy of the recordings. Table 2 also lists the mean periods of individual crests, taken as the time interval between adjacent troughs,

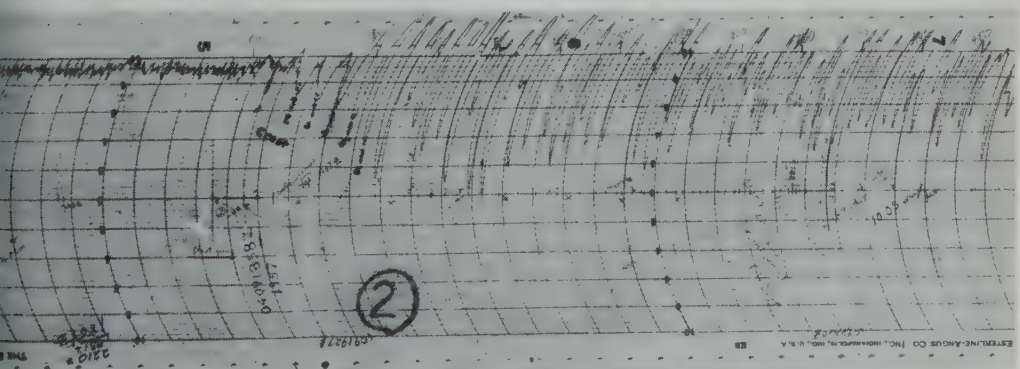


Fig. 8. Wave record of early portion of tsunami of March 9, 1957, at Wake Island. Time scale, 10 min/division.

TABLE 2. Wave Arrival Times and Mean Period at Island Stations
(Times are in minutes after initial disturbance; periods are in minutes.)

Station	Event	(1)		(2)		(3)		(4)		(5)		Crest
		Crest	Trough	Crest	Trough	Crest	Trough	Crest	Trough	Crest	Trough	
Ailinginae	Cherokee	11.1	13.5	15.0				Later arrivals questionable				
(±2 mm)	Zuni	11.1	14.0	15.7	17.0	18.3	19.4	20.4	21.0	21.8	22.6	23.3
	Navaho	11.7	14.0	15.5	17.5	18.5	20.2	21.3	23.5	25.0		
Period	(Zuni)	5.8		3.0		2.4		1.8		1.6		1.4
Eniwetok	Cherokee	34.0	36.3	40.2	43.0	46.0	48.2	50.3	53.0	56.5	58.5	61.5
	Zuni	34.5	37.1	40.2	43.1	45.4	48.3	50.7	53.0	56.0	59.5	62.1
	Navaho	34.0	37.2	40.2	43.2	45.9	48.2	50.2	52.8	56.6	58.7	60.5
	Tewa	34.0	37.3	40.0	42.7	45.6	47.7	50.0	52.3	55.0	57.5	60.5
Mean		34.1	37.0	40.2	43.0	45.7	48.1	50.3	52.8	55.9	58.5	61.1
Period		5.8		6.1		5.1		4.7		5.7		4.7
Wake	Navaho	69.5	73.1	77.5	77.8	83.5	85.0	88.0	90.0	92.0	94.0	95.5
	Tewa	69.0	73.0	76.0	79.5	82.7	85.5	87.2	89.5	91.0	93.0	94.5
Mean		69.3	73.1	76.8	79.7	83.1	85.3	87.6	89.8	91.5	93.5	95.0
Period		7.6		6.6		5.6		4.5		3.8		3.0
Johnston	Navaho	219.5	225.0	230.5	235.0	239.5	242.0	246.5	250.0	253.0	256.0	258.5
	Tewa		225.0	230.5	234.5	238.5	242.5	246.5	249.8	252.0	255.0	257.5
Mean		219.5	225.0	230.5	234.3	239.0	242.3	246.5	249.7	252.5	255.5	258.0
Period		11.0		9.3		8.0		7.4		5.8		4.8
		Trough	Crest	Trough	Crest	Trough	Crest	Trough	Crest	Trough	Crest	Trough
Wake	Tsunami	272.0	275.0	282.5	286.0	291.0	294.0	298.0	301.5	304.5	308.0	310.5
Period		12.0		10.0		8.0		7.5		6.5		5.5

except that for the leading crest the period was considered to be twice the time interval between this crest and the following trough.² No correction was applied for the phase, since this correction was always substantially less than 1 minute (see Fig. 3). Similar arrival times at Wake Island for the tsunami of March 9, 1957, are also given in Table 2. In this case the first recorded arrival was a trough.

Wave amplitudes. The relative amplitudes of the first six crests (where available) of a composite wave train, normalized to a common

TABLE 3. Corrected Relative Wave Amplitudes Normalized to Common Energy

Station	Crest Order Number					
	1	2	3	4	5	6
Ailinginae	4.6	3.9	3.1			
Eniwetok	3.8	3.2	3.0	3.8	2.6	2.4
Wake	0.49	0.76	1.0	0.64	0.58	0.43
Johnston	0.22	0.45	0.51	0.35	0.31	0.25

² The interpretation to be placed on the first wave period so determined is highly questionable (see section IV).

energy, are given in Table 3. This normalization was accomplished as follows. Because of excellent phase coherence between corresponding crests from different shots recorded at Eniwetok, amplitude differences were construed to be solely to differences in the source energy geometry of the respective shots. Navaho and Tewa shots had nearly identical source geometries and essentially equal amplitudes at Eniwetok, and they were therefore presumed to have equal energies. These two shots also provided the only usable records at Wake and Johnston. Therefore, the recorded wave amplitudes at these two stations were taken to be one-half the number of chart divisions between each crest and the succeeding trough. The amplitudes of these two shots, were averaged together to give the average amplitude of the three islands.

At Ailinginae, only Zuni shot had a direct exposure toward the recording station. Shots from Navaho and Cherokee having traveled a circuitous route around both Bikini and Ailinginae. The waves from Zuni shot were considered to be the more representative of the island as amplitude and phase were concerned. To compare the amplitudes of individual waves at Ailinginae with those from Navaho and Cherokee at the other three stations, the former wave

the (roughly) constant amplitude ratio observed to exist between the same recorded waves from Zuni, and those from Tewa and Navaho, all recorded at Eniwetok.

The amplitudes so determined were then used for the gain setting of the respective amplifiers and the reciprocal of the response used for the hydraulic filter (Fig. 3) to obtain the best estimate of the relative amplitudes recorded at these four stations. The response is a function of wave period, the relative values being taken from Table 2. It is difficult to place a reliable estimate on the probable errors inherent in the wave amplitudes reported in Table 3. The wave recorders were calibrated daily, except for Ailinginae at which the recorder was calibrated only at the beginning of each record. In the laboratory, the calibration of these recorders with a constant-frequency, constant-amplitude input indicates that the absolute accuracy of the system is about 0.5 chart divisions (2 per cent of the signal), although the relative accuracy for perturbations was about 0.1 chart divisions. The percentage accuracy of the recordings is thus worst for the small signals, and the absolute accuracy was probably not better than the square root of the random background noise. Fortunately, the background noise at Navaho and Tewa was extremely low, so that an error of 10 per cent of the reported amplitudes does not seem unwarrantable, and even a 20 per cent error would not materially alter the conclusions reached herein.

IV. ANALYSIS OF THE DATA

Experimental objectives. The object of this analysis was to reconstruct the best possible empirical model of the space-time history of a transient wave train emanating from a localized (centered) disturbance in water of approximately uniform depth, in order that the results might be compared with existing theoretical models generalized to the problem of tsunami and other impulsive disturbances.

Analysis procedure. The problem of comparing wave arrivals at several different stations was complicated by several factors. First, neither

the source nor the observation stations were in deep water, so that the water depths were not only nonuniform but differed on the average from station to station. Thus the dispersion history of the wave trains was in general different for each station, and it was necessary to ask whether, or how, the dispersion relationships could be expressed in parametric form for comparison of the respective amplitudes and phases of individual waves in a common coordinate system. Second, all the waves observed were transitional, being neither deep nor shallow water waves in the usual sense, so that their phase and group velocities depended on their wavelengths as well as on the water depth. Lastly, the amplitude-frequency spectrum of the source was not known either for the bomb waves or the tsunami.

Since there is no general theoretical treatment of a centered waves system in water of variable depth, the procedure used here was to approximate the actual depth variation $h(r)$ by an 'effective' depth h' , based on computed arrival time for infinitely long waves at the four observation stations. The pertinent variables were then expressed in terms of h' :

Distance

$$R = r/h'$$

Time

$$T = t/\sqrt{g/h'}$$

Wave number

$$\alpha = kh' = 2\pi h'/\lambda$$

Wave frequency

$$\gamma = 2\pi/T = 2\pi/\tau\sqrt{g/h'} \quad (1)$$

It is also assumed in the following that the wave behavior is everywhere governed by the linear theory of gravity waves, since the parameter $A\lambda^2/h'^3$ (amplitude \times wavelength²/depth³) proposed by Ursell [1953] is always much less than unity. Hence the following fundamental relationships obtain:

Wave velocity

$$C = dR/dT = c/\sqrt{gh'} = (\tanh \alpha/\alpha)^{1/2}$$

Group velocity

$$\begin{aligned} V &= R/T = v/\sqrt{gh'} \\ &= \frac{1}{2}C(1 + 2\alpha/\sinh 2\alpha) \end{aligned} \quad (2)$$

the values read from these records are, if any, minimal, since any inherent pen friction on the Angus recorder will reduce the extent of the pen excursions for a given signal.

TABLE 4. Station Distances, Effective Depths, and Travel Times over Least-Time Paths from Bikini to Island Stations

Station	h' , ft $\times 10^{-3}$	r , ft $\times 10^{-6}$	t_1 , min	Phase Lag, min	t_{1c} , min
Ailinginae (Zuni)	9.8	0.37	9.0	2.2	11.2
Eniwetok	12.2	1.13	30.0	3.3	33.3
Wake	15.4	2.81	65.0	4.8	69.8
Johnston	15.9	9.15	213.	7.0	220.
Wake (tsunami)	18.5	12.6	272.	0	

The approximation involved in assuming an effective depth in the above relations is discussed below.

Determination of the effective depth. For very long waves in water of variable depth $h(r)$ the travel time t to a station at range r from an initiating disturbance (considered small compared with r) is

$$t = r/\sqrt{gh} = g^{-1/2} \int_0^r h(r)^{-1/2} dr \tag{3}$$

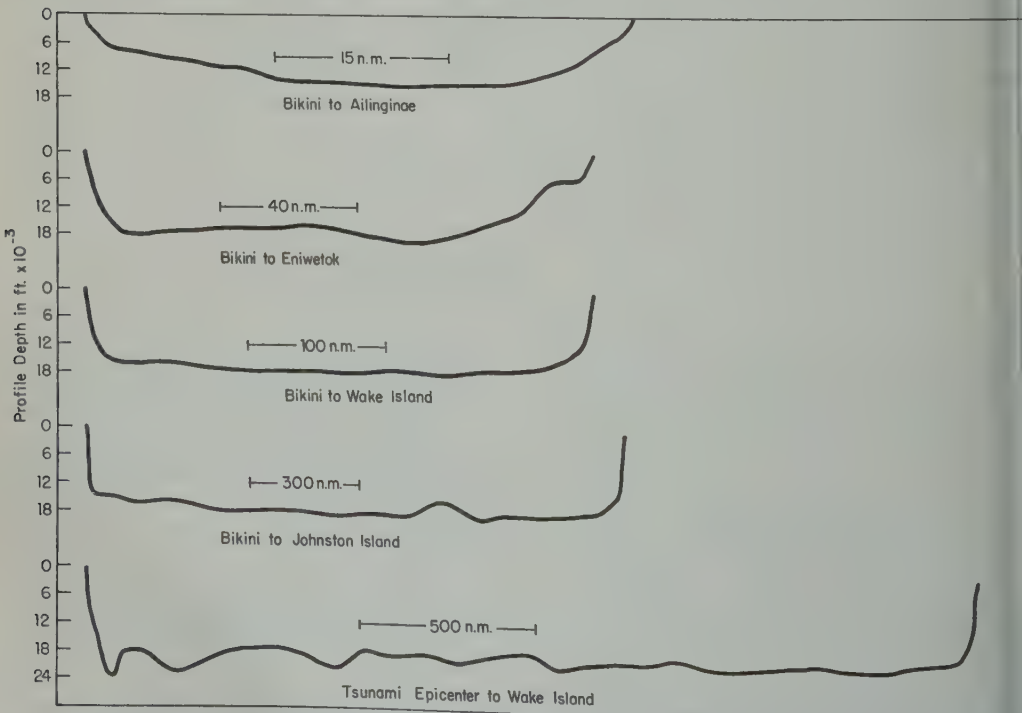


Fig. 9. Bathymetric profiles of the least-time paths from Bikini atoll to the four observation stations, and from the epicenter of the tsunami of March 9, 1957, to Wake Island.

and hence the effective depth h' is given approximately by

$$h'^{-1/2} \doteq \frac{1}{n} \sum_1^n h_i^{-1/2} \quad i = 1, 2, \dots, n$$

where the h_i refer to the local water depths assumed constant over small equal increments of r . A further restriction on t is that it be a minimum over the distance r , which, for complex bottom topography, is not in general along the most direct path between source and receiver. Determination of the minimum-time paths in the present case was accomplished by the graphical construction of successive wave fronts over 5-minute increments of travel time by Huygen's principle, according to the method of Hirono [1952]. The minimum time of arrival at each station was then determined by interpolation between fronts, and the minimum-time path was taken as that ray orthogonal to the fronts which passed through the station in question. The computed travel times t_1 , distances r , and effective depths h' so determined are tabulated in Table 4, and the bathymetric profiles along the minimum-time paths are shown in

Figure 9. The wave fronts for 20-minute time intervals from Bikini are also shown in Figure 2. A comparison of the observed and computed arrival times for the first bomb wave arrivals (Tables 2 and 4) reveals that the observed times were later than expected by an amount of time which increased with station distance, whereas the tsunami arrived essentially 'on time.' As will be shown later, however, wave generation theory predicts a phase lag for an impulsively generated wave train, as distinct from one originating from a localized deformation of the water surface. Since the bomb waves were clearly generated by the air impulse from the bomb explosions, these results indicate that the tsunami originated from an abrupt change in elevation of the sea floor and concomitant deformation at the free surface. The appropriate phase lags and corrected arrival times (t_{10}) are given in Table 4, and it can be seen that the latter are in good agreement with the observed values.

Theoretical considerations. Approximate solutions to the wave equation for centered waves in water of finite depth have been given by Eckart [1948] and Jeffreys and Jeffreys [1950] for one-dimensional plane waves and by Takahashi [1957] and Kranzer and Keller [1955] for cylindrical waves.

In the order listed, the dimensionless forms of these solutions are⁴

$$\sim A_0(2/T)^{1/3} A_i[(2/T)^{1/3}(R - T)] \quad (5)$$

$$\sim A_0 R^{-5/6} \cos(\gamma T - \alpha R - \pi/4) \quad (6)$$

$$\sim A_0 R^{-1} \cos(\gamma T - \alpha R) \quad (7)$$

where A_i is Airy's integral, and A_0 and A are the amplitude spectra at the source and point of observation, respectively.

It is advantageous to preface discussion of

These solutions apply to the case of initial deformation of the water surface, which is the case treated by Eckart and Jeffreys. The solution corresponding to (7) for the case of an initial impulse is identical to that given, except that cosine is replaced by sine, and was used in computing the phase lags for the first arrivals described above. In the following it will be assumed for convenience in comparing theory and results that the bomb waves originated from an initial deformation, the appropriate phase lag of the first wave being subtracted from the arrival times of all ensuing waves.

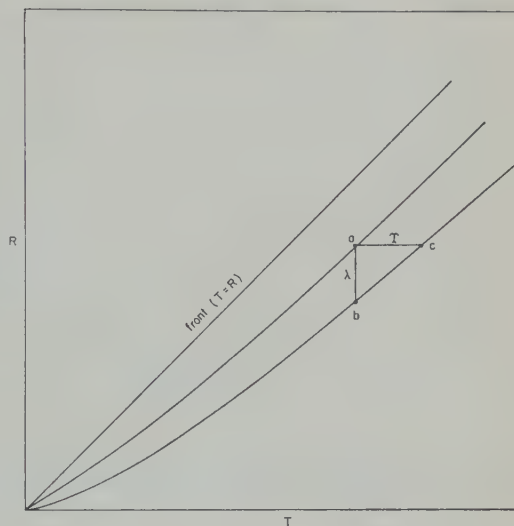


Fig. 10. Schematic R - T diagram showing the wave front and two successive wave phases emanating from a centered disturbance in water of uniform depth.

these solutions by a short qualitative description of some of the properties of a dispersive system of gravity waves originating from a localized disturbance in water of uniform depth. These properties are conveniently illustrated by considering the history of the development of such a wave train in a coordinate system having the dimensionless variables R and T as ordinate and abscissa, respectively, as shown in Figure 10. The expanding circular train is bounded by an initial crest or trough, depending on the sense of the original disturbance, which propagates outward at the velocity $C = dR/dT = 1$, as shown by the straight line $R = T$ in the figure. The space-time history of successive wave phases are curves in this presentation and represent solutions of the equation

$$dR/dT = C[\alpha(RT)] \quad (8)$$

of the form

$$(\gamma T - \alpha R) = \text{const.} \quad (9)$$

The variables R , T , α , and γ are related by the 'field' equations

$$\frac{\partial \alpha}{\partial T} + V \frac{\partial \alpha}{\partial R} = 0 \quad \text{and} \quad \frac{\partial \gamma}{\partial T} + V \frac{\partial \gamma}{\partial R} = 0 \quad (10)$$

which are derived very generally from the linear theory of dispersive waves [see, for example,

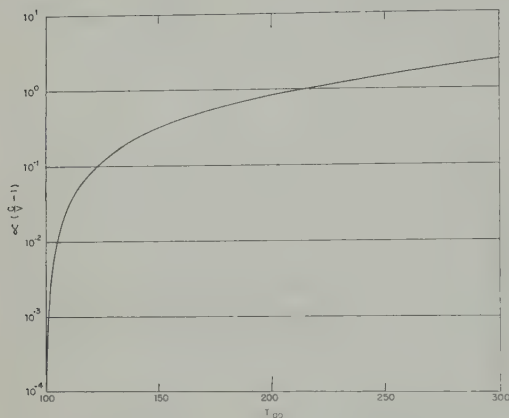


Fig. 11. Plot of the function $F(\alpha) = \alpha(C/V - 1)$ versus dimensionless arrival time T for the case $R = 100$.

Munk, 1947]. Equations 10 give the distribution of wavelength and period in the R - T plane, and for the special case of constant wavelength or period have solutions of the form

$$V = R/T = \text{const.} \quad (11)$$

The latter are straight lines, or rays, passing through the origin, the first (or earliest) of which coincides with the wave front, in view of the fact that $C = V$ at the front.

The vertical and horizontal separation of successive crests at any point on the R - T diagram (as shown by the intervals ab and ac in Fig. 10) have been defined as the wavelength and period, respectively, but both quantities are, in fact, scalar point functions in this representation, and the approximation becomes increasingly poorer near the front. The first visible intumescence (or depression) of an advancing train contains a large fraction of the total frequency spectrum, so that it is impossible to define its wavelength meaningfully except in the Fourier sense.

All the solutions to the wave equation mentioned in the first paragraph of this section have in common that the final result is obtained through the approximate evaluation of a Fourier integral by the method of stationary phase. The degree of approximation with this method is directly related to the extent to which the wave train is dispersed, or resolved [Eckart, 1948], and hence is poor close to the source. The ordinary quadratic approximation for the phase

of the integrand of the Fourier integral, as used by Kranzer and Keller, is also poor near the wave front. The other three authors have used a cubic approximation which results in a slight change in the dispersion but a significant increase in the predicted rate of amplitude decay with distance for the waves near the front.

Dispersion of a centered wave system. An important feature of all the foregoing theories is that the dispersion is independent of the nature of the source disturbance. Thus, the arrival times of successive wave phases observed at a distant observation station are expected to be the same (except for sense), irrespective of the amplitude spectrum at the source.

In terms of the dimensionless variables already defined, the solution of Eckart and Jeffreys predicts the positions of successive crests at a given time by the expression

$$R_i = T + a_i(T/2)^{1/3} \quad (12)$$

where the a_i are the positive turning values of the Airy integral [Miller, 1946], the first six of which, to three significant figures, are $-1.01, -4.82, -7.37, -9.53, -11.48$, and -13.32 . It is to be noted that successive crests are separated in space by a distance (wavelength) proportional to $T^{1/3}$. Since near the front $R \approx T$, a permutation of (12) giving the arrival times of crests at a fixed point is

$$T_i \approx R - a_i(T/2)^{1/3} \quad (13)$$

and the wave period is predicted to increase as $T^{1/3}$.

Crest arrivals at a fixed point are predicted by Kranzer and Keller—and by Takahashi—except for a term $(-\pi/4)$ in the argument of the cosine—whenever $\cos(\gamma T - \alpha R) = 0$; that is, when $(\gamma T - \alpha R) = 2\pi m$, $m = 0, 1, 2, \dots$. In view of the identities $T = R/\gamma$, $\gamma = \alpha C$, the last expression may be put in the form

$$2\pi m/R = F(\alpha) = \alpha(C/V - 1) \quad (14)$$

The solutions to this equation for the crest arrivals are most easily obtained graphically, making use of Figure 11, which is a plot of $F(\alpha)$ versus T for the case $R = 100$. This curve was constructed by assigning values to α for which C and V are determined from (2). The variables are plotted for convenient interpolation in Figure 12. Arrivals for other values of R

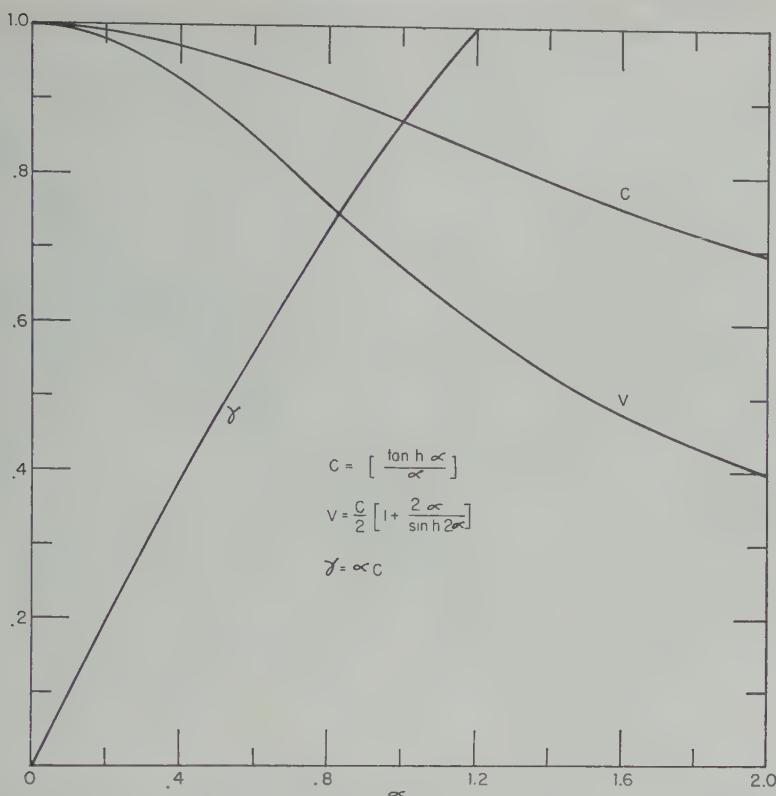


Fig. 12. Dimensionless phase velocity C , group velocity V , and wave frequency γ , as a function of wave number α .

obtained from Figure 11 by straight proportionality. The variation of wavelength with distance may be seen from the asymptotic expansion of (14), which, correct to third order is

$$\frac{\lambda}{R} = \alpha \left[\frac{\sinh 2\alpha - 2\alpha}{\sinh 2\alpha + 2\alpha} \right] \\ = \alpha \left[\frac{\alpha^3/3}{1 + \alpha^2/3} \right] \quad (15)$$

which it is apparent that $\alpha \sim R^{-1/3}$ for $\alpha \ll 1$, and $\alpha \sim R^{-1}$ for $\alpha \gg 1$. The first result is the same as that already obtained from the theory of Eckart and Jeffreys, and the latter is known from the linear theory for deep water waves.

A direct comparison can now be made between the predicted and observed dispersion, as evidenced by the arrival times of individual wave crests. Figure 13 is an R - T diagram showing 'trajectories' of the first six crests from an initial local elevation of the water surface,

constructed according to the theory of Kranzer and Keller. The vertical column of solid points gives the predicted positions of these same crests by the theory of Eckart and Jeffreys at the time $T = 681$, taken to coincide with the first arrival of the tsunami of March 9, 1957. Since this arrival was a trough, the presumption is that the initial displacement of the water was downwards, the sense of first motion being preserved throughout the history of the developing wave train. It is evident that the curves predict the observed arrivals with an error of less than 1 per cent and that the arrivals predicted by higher phase approximation of Eckart and Jeffreys are in even better agreement. One is led, therefore, to the conclusion that there is little significant difference between the dispersion predicted by the two methods discussed. This result may appear surprising, since the crest trajectories drawn are in fact curves of constant (stationary) phase θ , as given by the expression

$$\theta = (\gamma T - \alpha R) = 2\pi(n - 1) \quad (16)$$

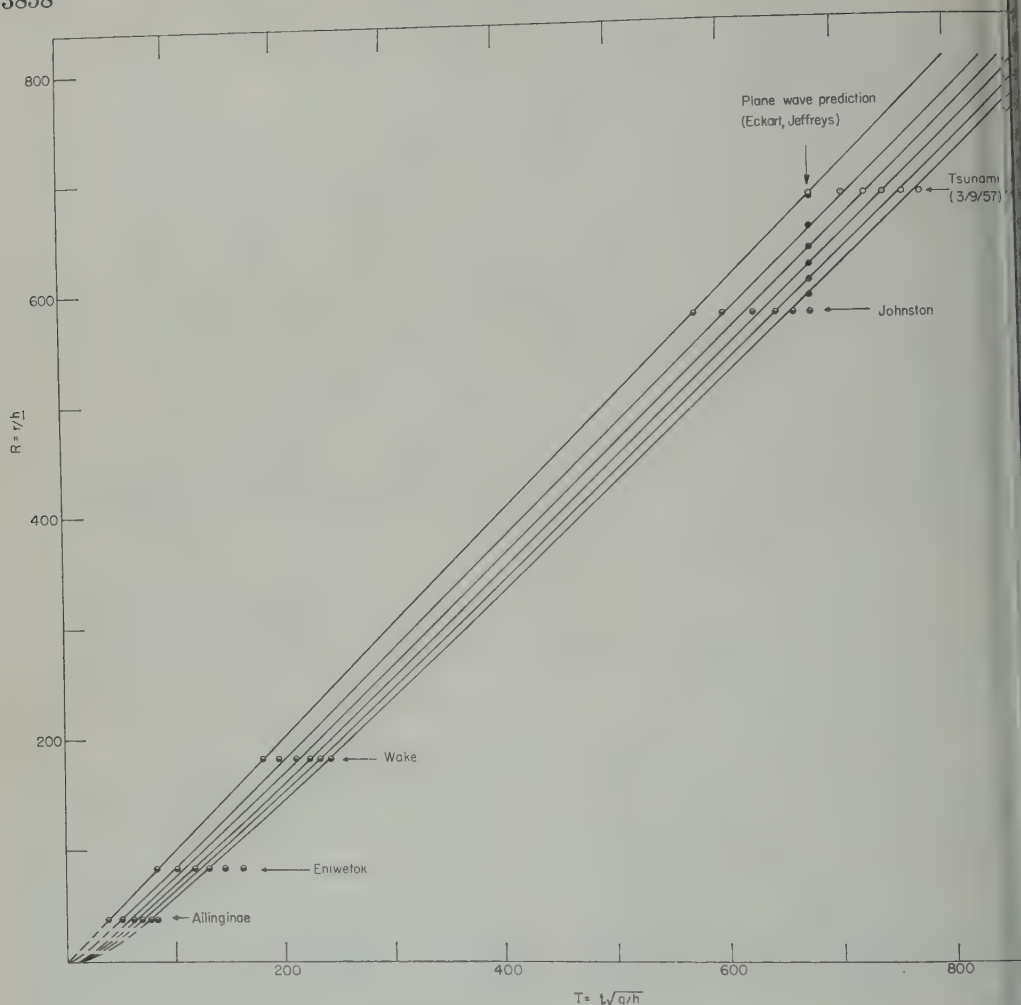


Fig. 13. R - T diagram showing first six crest arrivals predicted by theory of Kranzer and Keller. Data points give predictions by theory of Eckart and Jeffreys and observations of arrivals at island stations for same six crests for bomb waves and the tsunami of March 9, 1957.

where $n = 0, 1, 2, \dots$ is the order number of the crest arrivals. One ordinarily thinks of a point of stationary phase as being a useful approximation only when the phase is oscillating rapidly ($n \gg 1$) in an interval over which the wave number α has sensibly no variation. As pointed out by Liebermann [1956], however, the stationary phase method sometimes yields accurate results where the ordinary criteria for its use do not apply.

Figure 13 also shows the first six observed bomb wave arrivals at the several island stations. The data points for the first two or three crests fall fairly close to the predicted curves,

but fall progressively later with increasing n . Of the four stations, agreement is best at Wake and poorest at Eniwetok. The later Eniwetok arrivals appear to be essentially nondispersive, the observed wave period even tending to increase with time after the third arrival. A strong and persistent background oscillation having essentially the same period (300 seconds) was frequently observed in the day-to-day records from this station, suggesting that the arrivals may have been simple, locally excited resonance phenomena.

In view of the good agreement obtained

on the observed and computed arrivals for tsunamis at Wake Island, the discrepancies in the case of the bomb waves appear too large to be accounted for by differences in the dispersion resulting from inaccuracy in the stationary wave approximations considered above. Two possibilities suggest themselves. First, the effect of backscattering from the islands upon which the recording stations were located, heretofore neglected, may have been important in producing a phase lag. This effect would be expected to become more significant at higher frequencies, as observed. Second, where the depth variations are very large over the travel paths, it may not be reasonable to compare arrivals at different stations in a common coordinate system based on an effective depth. The first possibility appears unlikely, since from acoustic scattering theory the maximum phase shift for (approximately) sinusoidal waves is $\pi/2$, while the observed shift is of order 2π . The second possibility is difficult, if not impossible, to answer even

qualitatively, but it can be seen from Figure 13 that very large variations in the effective depth produce only relatively small changes in the dispersion. This is because the position of any point on the R - T plane shifts along a curve of the type $R = mT^2$ as the depth is considered to vary, and these curves are, in general, quite similar to the curves of constant phase. It is presumptuous, perhaps, to expect better general agreement between theory and observations in view of the nonideal nature of these experiments and the fact that, except for the later arrivals at Eniwetok, the greatest discrepancies between predicted and observed arrival times was only 1 or 2 per cent.

The wave amplitude spectrum. The amplitudes of individual waves at any instant of time depend not only upon the dispersion but also upon the amplitude spectrum of the source disturbance and the increase in crest length due to geometric spreading. The influence of the source amplitude spectrum manifests itself in

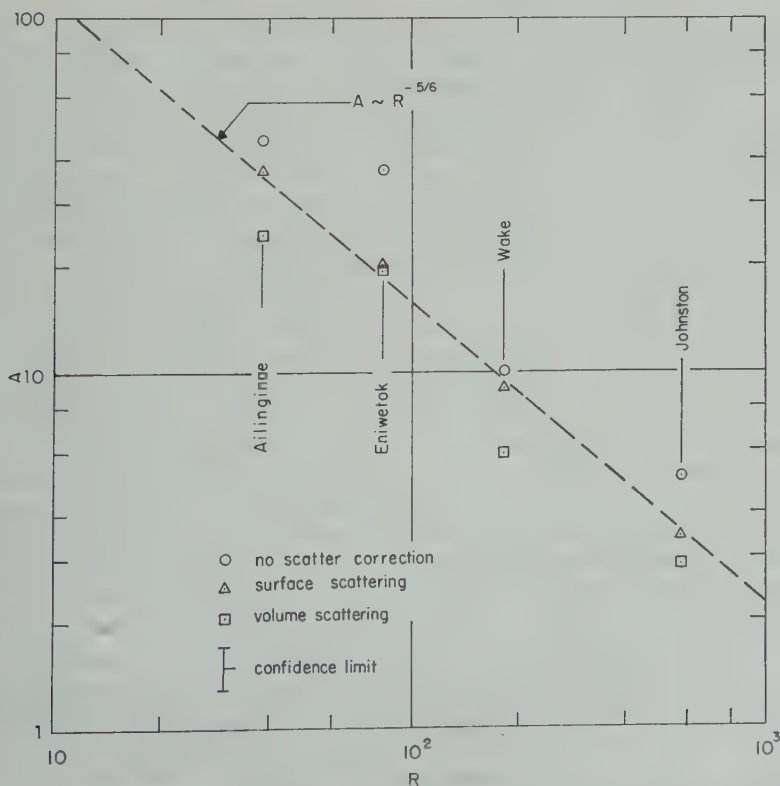


Fig. 14. Relative amplitude of highest observed wave at each station versus dimensionless distance.

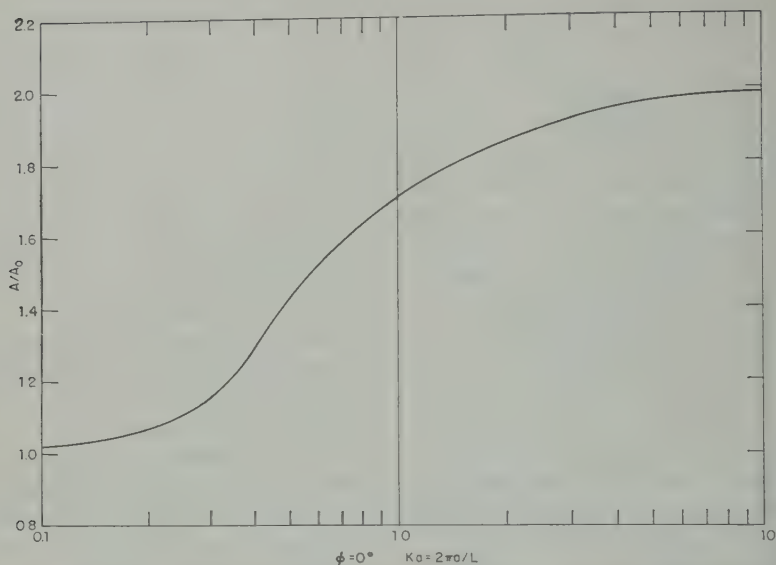


Fig. 15. Wave scattering enhancement as a function of island radius and wavelength for waves approaching a circular island at normal incidence, as predicted by acoustic theory.

a modulation of the dispersive wave train into one or more wave-envelope maxima (or groups) separated by nodes of (essentially) zero amplitude. In the development of Kranzer and Keller, and that of Takahasi, the nodes of the modulation envelope correspond to the zeros of the Bessel function $J_1(\alpha_0 R_0)$, where R_0 is a measure of the dimensions of the initial disturbance and α_0 is that value of α for which the phase is stationary; that is, where $V(\alpha_0) = R/T$. Evidently, since the first zero of J_1 occurs when $\alpha_0 R_0 = 3.83$, if the arrival time of the first node of the wave record can be clearly distinguished, the size of the initial disturbance can be determined. In the theory of Eckart and Jeffreys there is only one maximum of the amplitude envelope, the first crest is the highest, and it is followed by successively shorter and lower crests, in accordance with the oscillations of the Airy integral. This higher approximation, however, is valid only in the vicinity of the wave front, and it is to be expected that the ensuing waves should conform to the foregoing pattern after the first few crests have passed. The distinction between these two approximations is that the first predicts that the amplitude of the wave envelope at all points diminishes directly as the distance R from the source, while the latter pre-

dicts a decay⁵ proportional to $R^{5/6}$. In Figure 14 the circular points give the relative amplitude A of the highest wave observed at each station plotted against R , as determined from the experimental data of Table 3. These amplitudes, considered to represent the best estimates of the first wave envelope maximum, are not collinear in the log-log plot, although there is a general decrease in amplitude with increasing distance from the source. As with dispersion, it is possible that the discrepancy between the predicted decay law may be due to local enhancement of wave amplitude through scattering occasioned by the presence of the islands. It is to be expected that any effect of scattering would increase with island size, and the distribution of points suggests that collinearity would be improved by any correction having the opposite effect. Laboratory experiments by Wiener [1955] have shown that water waves are scattered by a vertical cylinder in a manner qualitatively predicted by acoustic theory, and a rough estimate can be made of the effect of scattering by small islands. Accordingly, the curve of Figure 15 was prepared from tabulated data by Wiener [1947], supplemented for small values of the product ka , by the asymptotic

⁵ A factor $R^{-1/2}$ applies to the plane-wave prediction to account for geometric spreading.

TABLE 5. Wave Amplitude Corrections for Island Scattering

Station	Maximum Amplitude	Local Depth h , ft	Wavelength λ , ft $\times 10^{-6}$	Island Radius		Scatter Correction	
				(Surface) a_s , ft $\times 10^{-4}$	(Volume) a_v , ft $\times 10^{-4}$	(Surface)	(Volume)
Maui	4.6	11,600	1.6	1.6	4.5	1.2	1.8
Okla	1.9	12,200	2.2	2.2	8.7	1.8	2.5
	1.0	17,900	3.1	3.1	5.4	1.1	1.1
Hon	0.51	18,000	3.7	3.7	7.2	1.5	1.2

ing formula given by Morse [1948]. 15 gives the ratio of the local wave-pres- intensity on the incident side of a cylinder us a to that at infinity as a function of ere k is the wave number at infinity. If ing theory has application to the case of indrical islands, a question arises as to the priate value of a . On the presumption that ost reasonable value should lie between extremes, the surface radius a_s and an ef- e radius $a_v = (Q/\pi h)^{1/2}$, where Q is the island volume and h the local water depth, ations were carried out for both cases. ch station the wavelength of the highest was calculated from the observed period e local water depth offshore. The results se calculations are given in Table 5 and rrected wave amplitudes are shown in e 14. The triangular points giving the wave udes corrected for surface scattering are ably collinear. Moreover the dashed line is the least-square mean of the four as has a slope $m = -5/6$, as predicted by analyses of Takahasi, Eckart, and Jeffreys. uare points giving the scattering correc- ased upon island volume show considera- read, although not as great as for the rrected data. It should be noted that the et expressing the confidence limits (± 20 ent) arising from the uncertainties in de- ing the original amplitudes is large enough at the possibility of a decay proportional cannot be ruled out, but it is felt that the r interpretation is both more reasonable ore convincing.

V. CONCLUSIONS

results of measurements reported here r to verify the predictions of linear theory

adequately, both as to the nature of the dis- persion and the rate of amplitude decay with distance, although the accuracy of the data is not good enough to permit a clear distinction between the quadratic and cubic approximation theories. The technique of predicting wave ar- rivals on the basis of an effective depth deter- mined from integration of the bottom profile is sufficiently accurate for practical purposes and should find general use in tsunami problems. In this connection it is of interest that the wave periods associated with the early waves from tsunami-like disturbances are rather smaller than was heretofore inferred from the study of tide-gage records, which suggests that the oscil- lations normally recorded on such gages are not tsunami waves at all, but simply the complex response of local water masses to the normal dispersion spectrum. This point of view is sup- ported by Munk, Snodgrass, and Tucker [1959], who point out that the spectra of tide records from the same station for different tsunamis are virtually identical, whereas those for the same tsunami at different gages are markedly dis- similar.

Since it is now reasonable to assume that dis- persion proceeds by linear processes, it should be possible to examine the motion at the source by constructing the appropriate Fourier trans- form from records made at a distance [Sató, 1956] or, conversely, by assuming various initial conditions and performing the indicated Fourier summations on a computer for comparison with observed wave forms. Either of these methods should lead to an understanding of the mecha- nism of tsunami generation, and both are now under study.

Acknowledgments. I am indebted to Professors W. H. Munk, Carl Eckart, and C. S. Cox, for re-

viewing this manuscript, as well as for their many helpful comments and suggestions.

This research was supported by contract Nonr 233(35) with the Office of Naval Research and the Defense Atomic Support Agency, and by contract NSFG 14104 with the National Science Foundation.

REFERENCES

- Eckart, C., The approximate solution of one-dimensional wave equations, *Revs. Modern Phys.*, 20(2), 399-417, 1948.
- Griggs, D. T., and F. Press, Probing the earth with nuclear explosions, *Research Mem. RM-2456-AEC*, Rand Corporation, Santa Monica, Calif., Sept. 1959.
- Hirono, T., and S. Hisamoto, A method of drawing the wave fronts of a tsunami on a chart, *Geophys. Mag. Tokyo*, 23(4), 399-406, 1952.
- Jeffreys, H., and B. S. Jeffreys, *Methods of Mathematical Physics*, Cambridge Press, 2d ed., p. 517, 1950.
- Kranzer, H. C., and J. B. Keller, Water waves produced by explosions, *Inst. Math. Sci., N. Y. Univ. Rept. 1MM-NYU 222*, Sept. 1955.
- Laird, A. D. K., A model study of wave action on a cylindrical island, *Trans. Am. Geophys. Union*, 36(2), 270-285, 1955.
- Liebermann, L. N., Extremely low-frequency electromagnetic waves, *J. Appl. Phys.*, 27(12), 1473-1483, 1956.
- Miller, J. C. P., The Airy integral, *Math. Part-vol. B.*, B.A.A.S. Cambridge Univ. Press, p. B43, 1946.
- Morse, Philip M., *Vibration and Sound*, McGraw-Hill Book Co., New York, p. 352, 2d ed., 1948.
- Munk, W. H., Increase in the period of waves traveling over large distances, with application to tsunamis, swell, and seismic surface waves, *Trans. Am. Geophys. Union*, 28(2), 198-211, 1947.
- Munk, W. H., F. E. Snodgrass, and M. S. Torrey, Spectra of low-frequency ocean waves, *Scripps Inst. of Oceanog.*, 10(4), 323, Univ. of California Press, 1949.
- Satô, Y., Analysis of dispersed surface waves by means of the Fourier transform, 2, Synthesis of the movement near the origin, *Bull. Earthquake Research Inst. Tokyo Univ.*, 34, 9-18, 1958.
- Takahasi, R., Model experiment on tsunami generation, Earthquake Research by Inst. of Earthquake Univ., unpublished, Sept. 1957.
- Ursell, F., The long wave paradox in the theory of gravity waves, *Proc. Cambridge Phil. Soc.*, 49, 685-694, 1953.
- Van Dorn, W. G., A new long-period wave recorder, *J. Geophys. Research*, 65, 1009-1016, 1960.
- Wiener, F. M., Sound diffraction by rigid spheres and cylinders, *J. Acoust. Soc. Am.*, 19(3), 1947.

(Manuscript received July 28, 1961.)

A Telemetering Ocean-Bottom Seismograph¹

JOHN EWING AND MAURICE EWING

*Lamont Geological Observatory, Columbia University
Palisades, New York*

Abstract. Successful tests of a telemetering ocean-bottom seismograph have been made on three occasions. In all cases, the seismograph was resting on the ocean bottom or planted in the sediments, sending its information to the surface by frequency modulation of a supersonic beam. The use of cables connecting the instrument on the bottom to the recording ship was avoided so that the level of background noise would not be influenced by shaking the instrument by a long cable. These first tests were designed to help determine what frequencies should be recorded, at what levels, and what method is best for transmitting to the recording instrument at the surface the earth's vibration in these frequencies. Neither the instrument used nor the method was the optimum for obtaining all of the seismological data from the ocean bottom, but they have demonstrated the feasibility of ocean-bottom seismographs and have helped to determine the criteria for the more complicated instruments and methods of transmittal which will ultimately make up a world-wide system. Data from such a system are expected to settle the question of the origin and propagation of microseisms, provide detailed information about the sedimentary layer and about the earth's crust and upper mantle, and, most important of all, may greatly increase the radius over which a single station can monitor small earthquakes and explosions. This will materially increase our ability to investigate the seismicity of the entire earth and also to monitor nuclear explosions.

In the preliminary tests body waves from one earthquake and several seismic refraction profiles were recorded. The earthquake record indicates reasonably good signal-to-noise ratio in the short-period range. The refraction profiles give indications from *P* and *S* waves of important regional and local variations in the character of the crust-mantle interface.

INTRODUCTION

This paper describes the history and present status of the development of seismographs to record earth vibrations on the ocean floor. The installation of seismographs on the ocean floor offers a number of attractive possibilities for investigations. Since much of the seismic background noise registered at land stations is attributed to meteorological and man-made causes, it is reasonable to expect that the deep ocean floor, far from the coast, might be exceptionally quiet. If so, we might hope to operate instruments of far greater sensitivity than those heretofore, thereby increasing our ability to monitor small earthquakes or large test explosions at much greater distances than at present. Fundamental studies of seismicity, which require that data on magnitude, depth, origin, focal mechanism, etc., be as nearly complete and uniform as possible, and for shocks

as small as possible, will be greatly facilitated by the use of more sensitive seismographs as well as by a more uniformly spaced world-wide network of receiving stations, requirements which may readily be met if ocean-bottom stations are available. The effectiveness of world-wide monitoring for large explosions would likewise be increased.

With stations on the ocean floor it will be possible to determine whether there are waves whose propagation is confined to oceanic structure and cannot be recorded at land stations because of distortion or attenuation at the continent-ocean boundaries. Pertinent data about the character of the boundary and information on oceanic structure which can be derived from studying earthquake waves having entirely oceanic paths may be obtained. At present, observations on islands offer the nearest approach to this capability, and they are limited by noise level and by the effects of the island structure itself on the propagation of waves.

In explosion seismology, receiving stations on

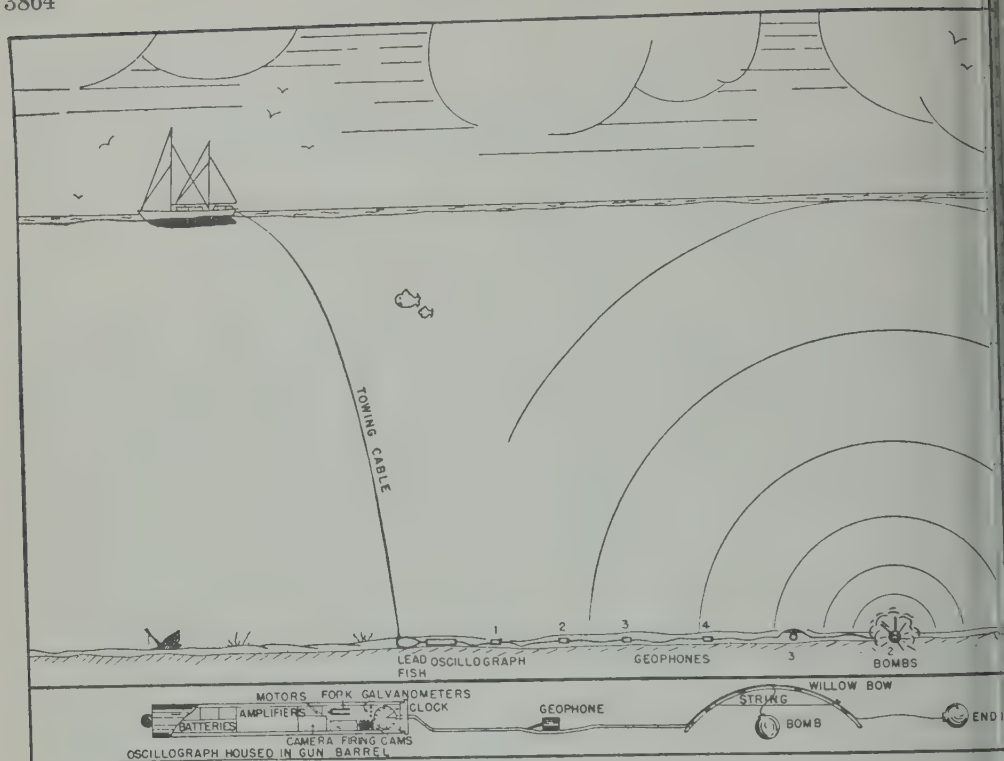


Fig. 1. Cable method of making deep seismic measurements.

the ocean floor will permit a more precise study of all the outer layers of the earth, including the sediments, crust, and upper mantle.

HISTORY OF THE DEVELOPMENT OF THE OCEAN-BOTTOM SEISMOGRAPH

In 1937 and 1938 seismographs were put on the ocean floor at depths exceeding 2000 meters in several attempts to shoot seismic refraction profiles. Only one of these tests, in 3000-m depth, about 350 km south of Newport, R. I., gave data about ocean-bottom layering. In these experiments an automatic oscillograph, four geophones, and four bombs, all distributed along an electrical cable about 1 km long (as shown in Fig. 1) were lowered to the bottom on the end of a heavy steel wire rope, laid out to form a straight line, and left undisturbed for about 15 minutes while the profile was shot and recorded automatically [Ewing and Vine, 1938; Ewing, Woollard, Vine, and Worzel, 1946].

In 1939 and 1940 a new system was developed; seismographs and bombs, each containing a timer to make it function at the correct time,

were sent to the bottom in deep water, according to the plan illustrated in Figure 2, a small amount of seismic refraction data was obtained at two stations, one 250 km south of Cape Cod, Massachusetts, in 2600 m, and another 550 km northwest from Bermuda in 4000 m [Ewing, Woollard, Vine, and Worzel, 1946]. These instruments were allowed to drop to the bottom under ballast which was detached automatically to allow them to float to the surface after the tests were completed.

At this time the program was interrupted by World War II. When work was resumed in 1947, the availability of wartime apparatus and techniques, the information from earlier tests that the sediment layer was at least several hundred meters thick, plus a desire to obtain long profiles in order to investigate the thickness of the oceanic crust led to the adoption of a technique for using shots and receiving them at the sea surface. The effort to operate seismographs on the floor of the deep sea was postponed for several years.

The renewal of this effort was announced

at the seminar on microseisms held under the auspices of the Pontifical Academy of Sciences in response to the need for observations of the ocean floor to resolve some uncertainties about the generation and propagation of microseisms [Ewing and Press, 1952]. The program proceeded at a slow pace, owing primarily to lack of financial support; however tests and preliminary designs were made of seismographs which could be recovered after recording on the ocean floor for extended periods of time.

In 1958, after additional requirements for seismic recording on the ocean floor developed, Ewing and B. Lusk joined the project and it was decided that the best approach for the tests would be to telemeter the information to a surface ship by modulation of a supersonic signal sent out from the ocean-bottom seismograph.

Successful tests with the telemetering ocean-bottom seismograph were made in November 1959, in September 1960, and in January 1961. The problem of keeping proper position for the ship which received the telemetered signals

was greatly simplified by using on the second attempt a trainable transducer for receiving the supersonic signals from the bottom seismograph. In this way the direction toward the seismograph could be determined and corrections for the drift of the ship made before contact was lost. Continuous monitoring thus became possible under all conditions under which the receiving ship could maintain her position.

The unit now being used on the ocean bottom contains a short-to-medium-period vertical seismometer and an amplifier whose signals modulate the frequency of a 12-kc/s acoustical source that sends signals in a broad beam toward the sea surface. The battery is adequate to operate this unit for 7 or 8 days.

The signals received on the monitoring ship are recorded on magnetic tape. The signals from seismic refraction shots are also recorded as conventional seismograms, which include time markers and shot-instant signals received by radio. In some experiments the monitoring ship has also used a near-surface hydrophone to record, for comparison with the ocean bottom seis-

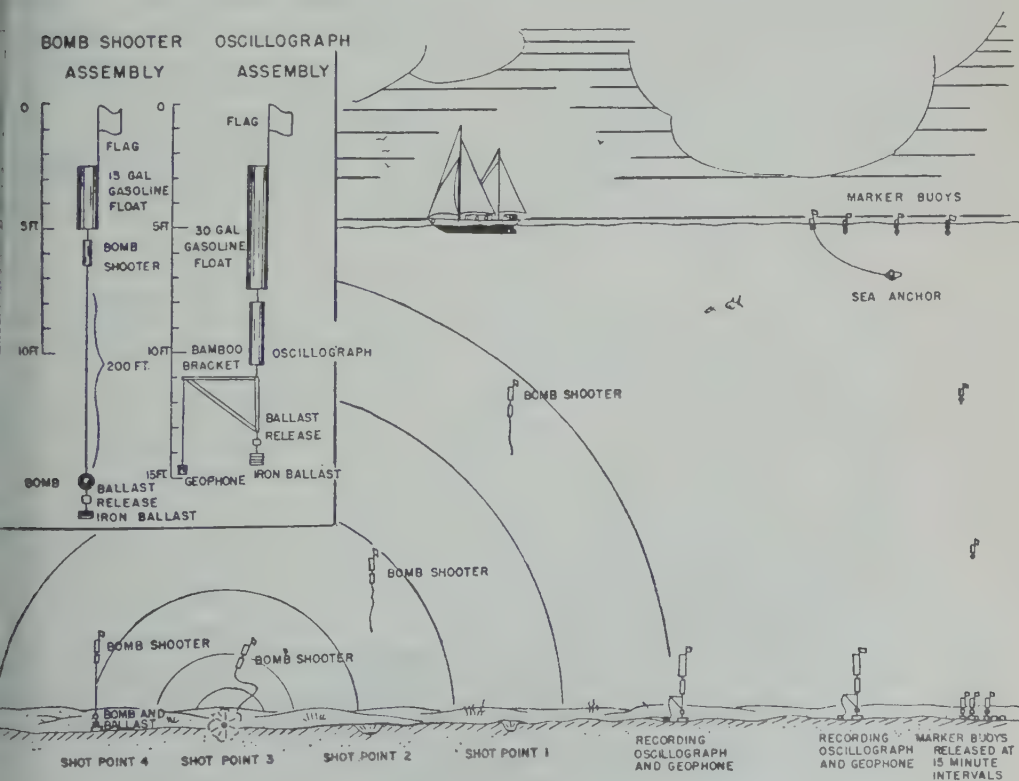


Fig. 2. Float method of making deep seismic measurements.

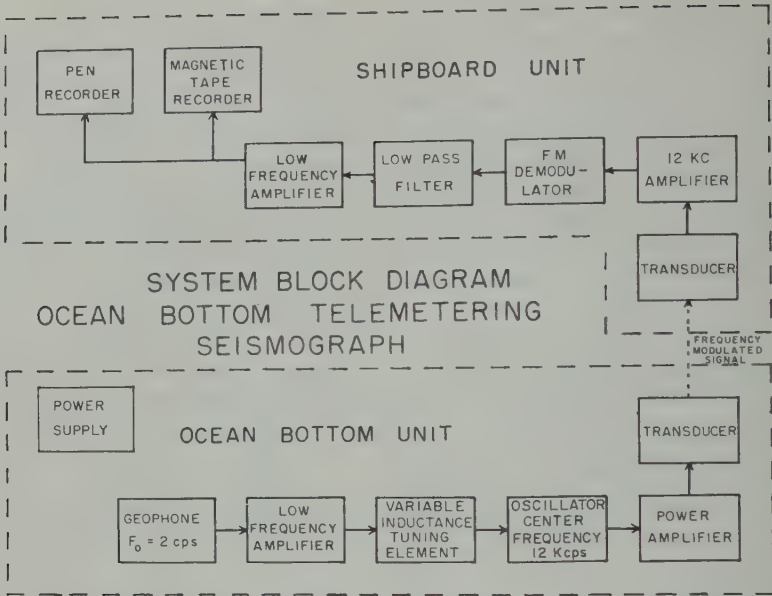


Fig. 3. System block diagram of telemetering ocean-bottom seismograph.

mograph data, seismic refraction data of the type which has been used extensively in our work during the past 15 years.

SOME DETAILS OF THE INSTRUMENTS USED

A block diagram of the telemetering ocean-bottom seismograph is shown in Figure 3. A geophone with natural period 2 cps is operated at about 0.5 critical damping as the detector of

the ocean-bottom unit. The signal from the geophone is amplified and used to drive a variable-inductance element which controls variations in frequency of the oscillator about 12 kc/s mean. The frequency-modulated output of the oscillator is amplified to deliver about 1 watt of power to the transducer.

The transducer is a small free-flooding piezoelectric unit, whose transmitting response

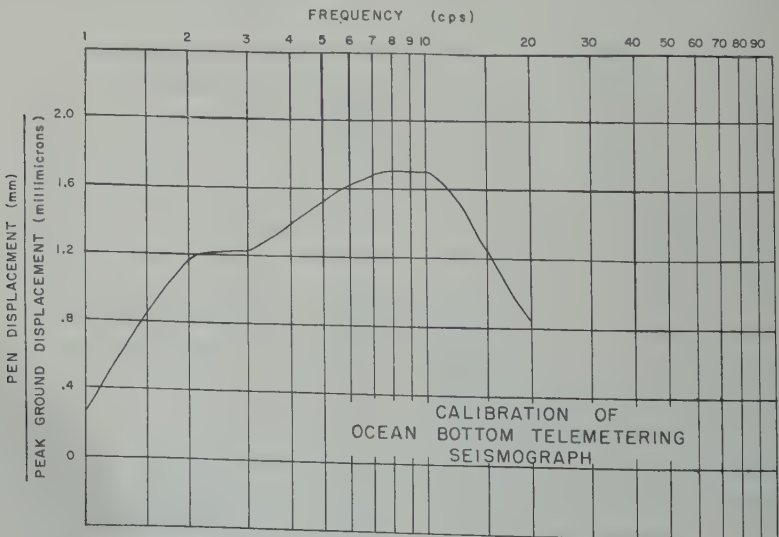


Fig. 4. Calibration of telemetering ocean-bottom seismograph.

TABLE 1. Bottom Seismograph Stations

Station	Date	Monitor Ship	Water Depth, m	Position		Azimuth Shot Line, deg	Length Shot Line, km
				Lat.	Long.		
	Oct. 10, 1959	<i>Vema</i>	4800	35°55.5'N	68°37.5'W	165	83
	Oct. 14, 1959	<i>Vema</i>	4995	35°02'N	66°43'W		
	Sept. 17, 1960	<i>Grace</i>	5000	30°17.5'N	65°55.5'W	000	52
						180	268
	Aug. 10, 1960	<i>Grace</i>	5000	30°18'N	65°55'W		
	Aug. 11, 1960	<i>Grace</i>	5000	30°18'N	65°55'W		
	Sept. 10, 1960	<i>Grace</i>	5000	30°18'N	65°55'W		
	Jan. 26-28, 1961	<i>Vema</i>	3700	24°00'N	91°27.5'W	090	235
						300	172
	June, 1961	<i>Vema</i>	5600	44°09'S	53°27'W		
	June, 1961	<i>Vema</i>	5600	44°27'S	53°35'W		

out 10 to 14 kc/s. The ocean-bottom unit is completely transistorized. Its power supply is a battery, estimated to provide for 7 to 8 days of continuous operation.

The telemetered acoustic signal is received on board with a standard UQN-1 echo-sounder transducer, which is tuned broadly to 12 kc/s. It is provided with a band width of about 2 kc/s. The signal is a beam width (between 10 db points) of about 45°. An amplifier covering the band 10 to 80 kc/s is used to amplify the signal from the transducer and to pass it into the FM demodulator (a commercial frequency meter). The output of the demodulator is filtered, amplified, and recorded on a pen recorder and on magnetic tape.

The FM telemetering system was chosen to permit a calibration of the entire system which would be independent of losses in the telemetering link. The calibration curve in Figure 4 was obtained by use of manufacturer's constants for the geophone and electrical measurements made in the circuits. The same calibration curve was used in several runs in the laboratory at temperatures from 27° to 0°C and in a run on shipboard just before the main unit was put into operation.

Three types of battery supply and launching methods were used. At station 1a (see Table 1) the unit was assembled on the framework of the deep sea coring equipment as shown in Figure 5, the geophone being in a pressure case at the bottom of a lead pipe about 3 m long and 7 cm in diameter. The signal leads from the geophone passed through pressure seals and then through

the pipe to a platform at the top. On the platform were the battery, the transducer, and a pressure case containing the amplifier and telemeter circuits. The lead-acid battery was insulated from sea water by rubber sheets which permitted pressure equalization. The transducer was free-flooding. The pipe was mounted beneath the 550-kg weight of the standard coring apparatus by means of a 'weak link' which would allow the weight to drive the lower part of the assembly into the sediment and leave it there. As in the standard coring operation, the assembly was lowered to about 5 m above bottom on the trawl wire, then released by the action of a bottom trigger weight, and allowed to fall freely, as illustrated in Figure 5.

At station 1, a second type of assembly was used, smaller and lighter than the first so that it could be lowered on the hydrographic wire. This apparatus, shown in Figure 6, was to be seated on the ocean floor, instead of being driven into it. The geophone was mounted on a cart-wheel base about 1 m in diameter to assist in maintaining an upright position and was used as the trigger weight in the coring apparatus. The transducer, the electronic circuits, and the battery, which consisted of 48 No. 6 dry cells potted in roofing tar, were placed in the coring head position. When the geophone was seated on bottom the trigger arm released the other package, allowing it to fall and seat itself separately.

At station 2, occupied in September 1960, a third type of bottom unit was used. This was a free-fall model, the design being dictated by the

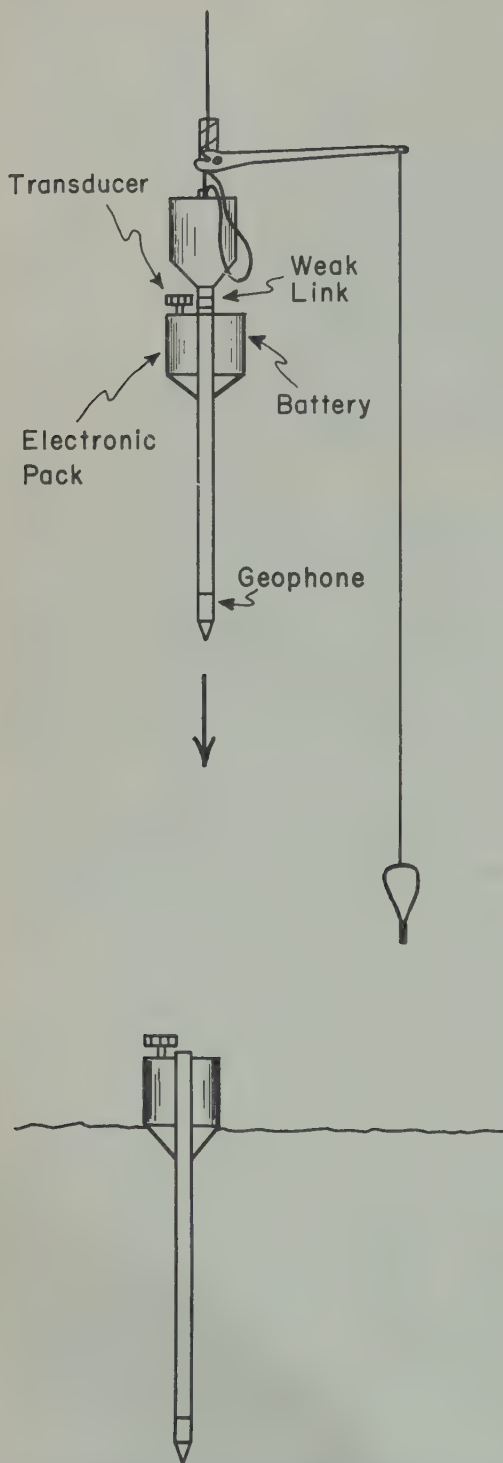


Fig. 5. Seismograph unit designed for placement on bottom with coring apparatus.

lack of a deep sea winch aboard. Figure 7 shows the arrangement of components and design of the 'missile.' The total length from the transducer to the tip of the nose pipe was about 30 m. Its weight in air was approximately 300 kg. This unit reached bottom in 10 minutes in the 4800 m deep. It struck bottom at a speed of about 8 m/sec without causing damage to any of the components.

At stations 3, 4a, and 4b, the standard pressure coring weight and release mechanism were used as the basic unit used in planting the seismograph. This unit was set gently on the bottom by release from the trawl wire. The ocean bottom unit which transducer amplifier and seismometer and hydrophone were mounted together, was attached below the core weight by a line about 20 m long. The trigger weight for release

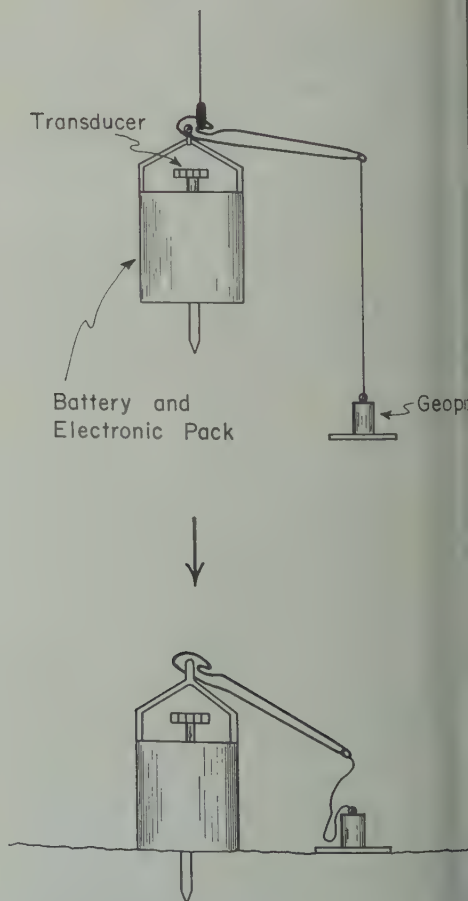


Fig. 6. Seismograph unit designed for placement on bottom with small-diameter hydrographic apparatus.

10 m below the release, or about 10 m above the unit, and displaced about 2 m laterally. It was possible to set the unit on the bottom by nudging contact by the telemetered signal, and to monitor it for proper operation for a few minutes with neither the main weight nor the trigger weight in contact with the bottom. If performance was satisfactory, the unit was released and held a few meters above bottom until a 'messenger' could be sent down the trawl wire to set the trigger mechanism in condition to operate. The trawl wire was then payed out until first the seismometer unit and then the trigger weight touched bottom. Contact of the trigger weight with the bottom allowed the main weight to drop about 1 meter and to sever the connections with the seismometer unit. Except for the method of placement on the bottom floor, the units described have been very similar functionally. The differences have been in minor details of electronic circuitry, power sources, or detector sensitivity. Each has used a short-period vertical seismometer modulating a radio carrier frequency. In the earlier part of the work, the transducer on the monitoring unit was rigidly attached to the hull, and the indication of the direction toward the telemetering unit was obtained by correlating the variation in carrier intensity with the roll of the ship. Later, provision was made for training and rotating the transducer in any desired direction, and the problem of maintaining contact with the unit on bottom was greatly simplified.

The demodulator system requires some improvement. On several occasions the carrier signal and the modulation of it by shots, could not be heard clearly, but were apparently below the threshold level for satisfactory operation of the demodulator recorder.

The system described here has many obvious serious disadvantages, some of which will be discussed. The system was, of course, never intended to serve for long-continued monitoring. For such monitoring, a cable connected to the unit or extended along bottom to such a depth that it could then be led to the surface without introducing mechanical noise in the seismograph would be one solution. Another solution would be to store the data at the seismograph, either on magnetic tape or on film, for periods up to possibly one month. Upon recovery from a monitoring vessel, the accumu-

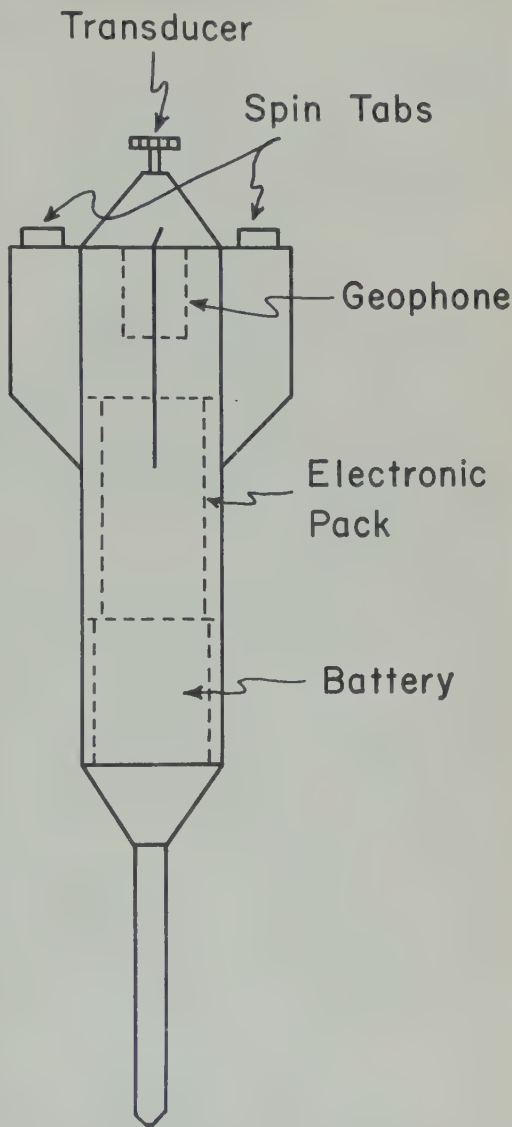


Fig. 7. Free fall ocean-bottom seismograph.

lated data could be delivered, either by accelerated telemetered playback or by transport of the record to the surface. Both of these systems are being investigated with a view to installing permanent or semipermanent stations. However, further studies with the instrument described here are essential in order to gain sufficient information about noise level, optimum sensitivity, etc., particularly in the longer-period region of the spectrum, so that the permanent installations can be properly designed.

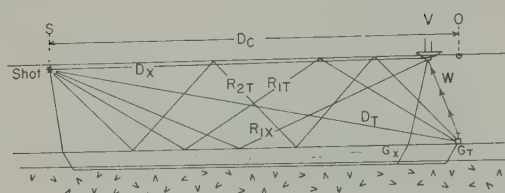


Fig. 8. Ray paths for water waves.

METHOD OF REDUCTION OF BOTTOM SEISMOGRAPH REFRACTION DATA

This method of operating and telemetering an ocean-bottom seismograph depends heavily on sound propagation through the water. All signals are received at the ship as acoustic signals, and these have traveled along a number of different paths (Fig. 8).

The signals G_T , D_T , R_{1T} , R_{2T} , . . . travel to the bottom seismograph, along the paths indicated, as conventional seismic and acoustic waves. From the seismograph they are transmitted to the ship as modulations of a 12-ke/s acoustical beam. First, it is necessary to determine the horizontal distance from the shot to the seismograph, or the equivalent quantity D , the time for the explosion sound to travel via the surface sound channel from the shot to a surface point directly above the seismograph. Second, it is necessary to determine the travel time W from the bottom seismograph to the ship, in order to correct observed arrival times for the delay introduced by telemetering.

The most direct way to determine W is to measure the interval between detonation of a small explosive charge near the ship and the return of a telemetered signal from that explosion, but in some situations this method is not available, owing to the unsuitable frequency response of the bottom seismograph or to blanketing of the telemetered signal by a strong echo from bottom.

There is also a comparison method for estimating W , which involves subtraction of W_1 , an approximate value, from all telemetered arrival times to obtain a first approximation to the true arrival times at the seismograph. The arrival times are then compared with those for paths such as G_T , D_T , R_{1T} , R_{2T} , . . . directly to a hydrophone near the ship, as shown in Figure 8. The method of comparison must be chosen in a particular way because the ray paths, which are

shown as straight lines, are actually curving owing to variations in velocity with depth.

From the paths shown in Figure 8 it may be seen that the following relations hold between travel times to the seismograph and those to hydrophone at distances chosen to keep inclinations of the comparable paths the same:

$$D_T = 1/2 R_1 (2D)$$

$$1/3 R_{1T} = 1/4 R_2 (4/3 D) = 1/2 R_1 (2/3 D)$$

$$1/5 R_{2T} = 1/6 R_3 (6/5 D)$$

$$= 1/4 R_2 (4/5 D) = 1/2 R_1 (2/5 D)$$

where D_T , R_{1T} , and R_{2T} are direct and reflected waves to the bottom seismograph, D is the direct wave to point O at the surface directly over the seismograph, and R_1 , R_2 , . . . are reflected waves arriving at that point. The expression $1/2 R_1 (2D)$, for example, signifies half the reflection time (surface-bottom-surface) for a ray at twice the range D . These equations are exact only for conditions of constant velocity with depth. Corrections are required for sloping and uneven topography.

At station 1, in the first long refraction profile recorded by a bottom seismograph (Fig. 9), the quantity W was not measured directly; hence it was necessary to use the water wave comparison method to determine shot-detector distances. This was accomplished as follows:

1. Water waves recorded by a hydrophone suspended near the surface from the receiving ship, yielding data for D_T , R_{1T} , R_{2T} , . . . were used to determine the curves D , R_1 , R_2 , and in Figure 9. The observed points are omitted to improve legibility.

2. From equations 1 and from the above curves, the relationship between direct and reflected travel times for arrivals at point O at the bottom seismograph were obtained. The curves are labeled D_T , R_{1T} , and R_{2T} in Figure 9.

3. Approximate values for D_T , R_{1T} , and the travel times to the seismograph, were determined for each shot by subtracting an amount $W_1 = 3.20$ sec from all telemetered water wave arrivals. This amount corresponds to the travel time from the bottom seismograph to a point on the surface approximately 1 km from point O. W_1 was chosen on the basis that an area of radius 1.5 km about point O is judged to be the area of detectability of the acoustic beam (1).

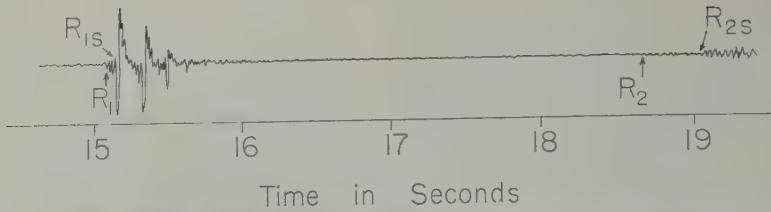


Fig. 10. Oscillogram showing bottom reflected and sub-bottom reflected arrivals.

on the directionality of the receiving transducer). The proper value of W would be between 3.30 and 3.15, the latter being appropriate if the receiving ship were at 0.

4. The resulting travel times (to the seismograph) were then used to determine a value for D from each shot, using equations 1 and the curves D , R_1 , R_2 , and R_3 . The values thus obtained are shown in Table 2 and compared with the values D_s , the direct water wave received at the ship.

5. The travel-time data for water waves and ground waves were plotted as squares in Figure 9, using the distances derived in this fashion, to permit comparison with the curves D_T , R_{1T} , R_{2T} and with the ground-wave data recorded by standard surface hydrophone technique (represented by triangles).

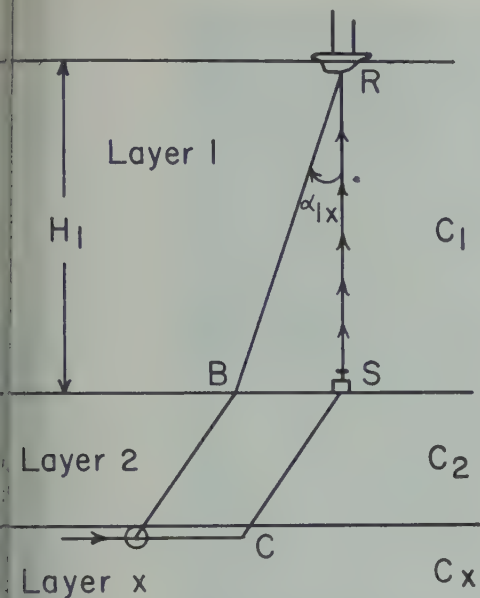
The agreement of the bottom seismograph data and those obtained in the conventional way is close, although there are systematic discrepancies in the water wave arrival times. These discrepancies indicate that the range from the shot to the receiving ship was less than that to the bottom seismograph for all shots. This suggests that the transducer on the bottom seismograph is somewhat directional and that the unit was slightly tilted. A similar transducer has been sent to a testing facility to determine the beam pattern, but the results have not yet been received. From the field observations we know that a sufficiently strong carrier signal for effective demodulation could be received only in a small area; hence even a small amount of directionality in the beam pattern could have produced the range discrepancies shown in Table 2. The fact that hydrophone-recorded ground waves plotted against hydrophone-recorded water waves produced essentially the same refraction lines as the equivalent quantities received by the bottom seismograph is a good indication that the method used to deter-

mine shot-to-seismograph distances was reasonably accurate. The errors in distances (times of the direct waves) between shots to the bottom seismograph are estimated to be ± 0.2 sec for ranges up to 35 km and no greater than ± 0.3 sec at the range of the longest shot.

The telemetered water wave arrivals in Figure 9 plot somewhat above the computed curve of D_T , R_{1T} , and R_{2T} . The fact that the R_{2T} points at short ranges are well above the curve is probably the result of poor high-frequency response in the bottom seismograph circuitry and sedimentary structure. As shown in the respective curve (Fig. 4), the sensitivity falls off sharply above 10 cps, and as shown in Figure 10 the tracing of a low-frequency hydrophone trace the strongest low-frequency arrival in the reflected wave is a sub-bottom reflection, labeled R_s , and corresponding to the points marked with small dots in Figure 9. Figure 10 shows reflection in which the angle of incidence on bottom is comparable to those for the later arrivals. The interval between the bottom

TABLE 2. Shot Distance Computations, Telemeter Station 1

Shot	D_s Measured at Vema	D , Computed from:		
		D_T	R_{1T}	R_{2T}
241	13.19		14.25	
242	16.97	17.92	17.95	
243	17.56	18.60	18.57	
244	18.25	19.39	19.35	19.77
246	19.17	20.39	20.24	20.65
247	24.85		25.40	25.58
248	25.67		26.15	26.40
249	32.02		32.46	32.50
250	33.96		34.65	34.64
251	39.67		40.53	40.45
252	40.31		41.31	41.23
253	44.77		45.44	45.35
254	50.07		53.40	53.48



11. Ray paths for arrivals to the surface directly and via the bottom seismograph.

tion and the sub-bottom reflection is comparable to the amount that the R_{2T} arrivals are. At the same range, owing to the difference angle of incidence of the rays on the bottom, sub-bottom reflected arrival in R_{1T} is almost coincident with the bottom reflection, and the observed arrivals fit the curve more closely.

The corrections required to reduce the telemetered ground-wave arrivals to the same surface of reference (sea level) as those recorded by the hydrophone are obtained as follows. Figure 11 shows the ray paths corresponding to a signal received via the bottom seismograph and one received at the surface directly by the instrument. If T_1 is the travel time along OBR and T_2 is that along $OCSR$, it is possible to show, with Snell's law, that $T_2 - T_1 = C_1 (1 - \cos \alpha_{ix})$, where $\sin \alpha_{ix} = C_1/C_x$.

In addition to the corrections for the surface reference and shot depth, all ground-wave arrival times (telemetered) were adjusted by 5 sec to account for the receiving ship's estimated horizontal distance from the seismograph.

PRELIMINARY RESULTS OF MEASUREMENTS

Seismic refraction profiles have been recorded in each of the three successful tests that have been made to date. Pertinent information about

them is given in Table 1. The data have been completely analyzed only for station 1. The results are discussed below. Also described are certain observations from stations 2 and 3 which were obvious in a less thorough examination of the data. Travel-time graphs are not shown for the last two stations but probably will appear in a later paper.

Station 1. Figure 12 shows tracings of three records from this station made with shot-detector distances of 25, 37, and 48 km. The charge size for shots 243 and 247 was 4 kg of TNT; shot 249 was 10 kg. Three ground-wave phases (P^* , R_1P^* , and mantle shear waves) are indicated and their correlation shown. The water waves indicated are the direct wave to the transducer on the ship's hull, the first-order bottom-reflected wave received at the transducer, the direct wave to the geophone, and the first- and second-order reflected waves to the geophone. The direct wave to the geophone was received only for the shots at ranges of about 35 km or less, owing to refraction in the water. Presumably, they would have persisted to somewhat longer ranges had the high-frequency response of the system been better. Figure 8 shows the ray paths for reflected and refracted waves received at the geophone and at the hydrophone and transducer.

The water waves which traveled directly or by reflection from the shots to the ship's transducer were identified by comparison with records made simultaneously with hydrophones floated near the surface. These signals represent only the portion of energy in the frequency range near 12 kc/s and appear as sharp impulses of short duration. They have little character other than a sharp beginning. The records for the water waves which travel to the geophone and then up to the transducer are sinusoidal if the amplifiers are not overloaded. Their characteristic frequency at the onset is in the neighborhood of 40 cps, the approximate upper cutoff frequency of the system. The frequency decreases with time owing to bottom penetration by the low-frequency sound.

The ground waves appear at frequencies between 5 and 15 cps. They are similar to waves received by hydrophones suspended near the surface. In the three records reproduced, ground motion associated with the first phases was about 10μ at 5 cps.

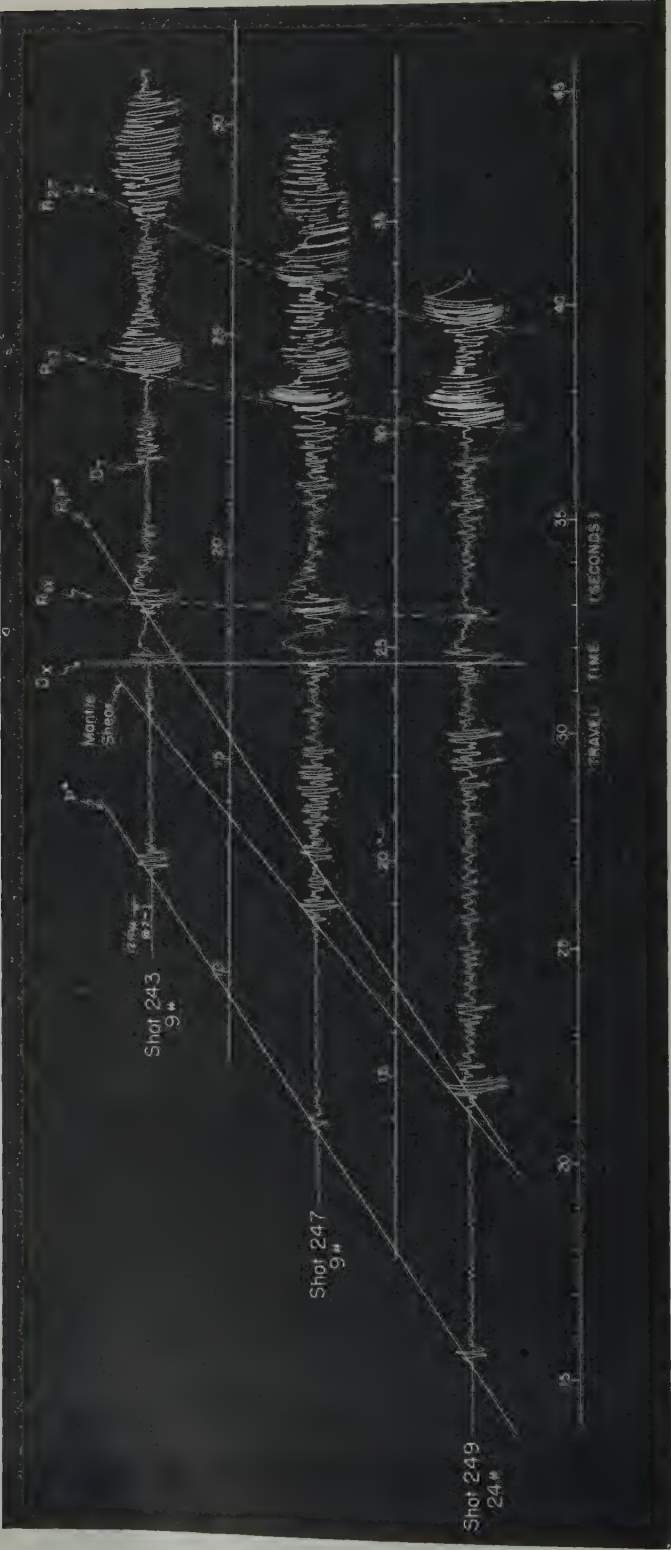


Fig. 12. Typical oscillograms from explosions, profile 1.

Figure 9 shows the time-distance graph and structure section deduced from the data. The metered arrivals were received at the same position as a short end-to-end standard profile (are plotted as the extension of one leg of triangles beyond 15 sec D time). Triangles denote arrivals at a suspended hydrophone; circles denote arrivals at the geophone on the bottom. Sub-bottom reflections are indicated by asterisks. All arrivals were corrected with sea level as the surface of reference.

The seismic velocity determined by each line is given (in km/sec) on the time-distance graph and in the structure section. In addition to direct and reflected sedimentary arrivals, compressional and transformed shear waves were observed for each of the three high-velocity layers (4.8, 6.6, and 8.1 km/sec compressional velocity). Shear waves propagating horizontally in each of these layers underwent transformation at the top of the 4.8-km/sec layer. Transformed shear waves in the 6.6-km/sec layer (β_4) have been observed frequently in standard marine refraction measurements [Herd, Officer, Johnson, and Bergstrom, 1952; Katz and Ewing, 1956; Ewing and Ewing, 1959] giving about 3.7 km/sec as an average shear velocity in this layer. Mantle shear waves have been recorded only rarely, and in general it has not been possible to determine an accurate velocity from the measurement. In the present experiment, distinct mantle shear wave arrivals were recorded by the geophone on the bottom over a range of 27.5 to 50 km. In some cases, this arrival was one of the most prominent on the record, e.g., shot 247 in Figure 12. It is not yet possible to say whether these arrivals were recorded strongly owing to some particular features of structure, or whether their reception was enhanced by receiving on the bottom rather than near the surface of the water. Transformed crustal shear waves are frequently received strongly at the surface, however, and both crustal and mantle arrivals cross the water-sediment interface as compressional waves. These facts tend to lessen the possibility that receiving on bottom was an important factor in the strength of the shear waves.

Regardless of the explanation of the strength of these arrivals, there is little doubt that they are mantle shear waves. The slope and intercept of the line through them is in general agreement

with that predicted by the structure derived from computations of compressional arrivals. The apparent velocity of mantle shear waves from this measurement is 4.3 km/sec. Although this is an unreversed measurement and may be affected by slope, indications from the compressional wave data are that any slope correction would lower, rather than raise, the shear velocity. The value of 4.3 km/sec is lower than that indicated by the study of earthquake surface waves [Dorman, Ewing, and Oliver, 1960]. It is difficult to attach any significance to this discrepancy on the basis of a single measurement; it will suffice here to point out that if there is a discrepancy it may be explained on the grounds that these seismic arrivals sampled only the uppermost part of the mantle where the velocity may be appreciably lower than at a slightly deeper level. It is certain that these arrivals did indeed sample the upper part, because critical distance for shear waves here is 26 km, and the first arrival was received at a range of 27.5 km.

Measurement of background noise at this station shows less than 1- μ displacement in the frequency range 2 to 10 cps. This agrees qualitatively with the results given by Brune and Oliver [1959] from a study of seismic noise at land stations. They found minimum displacements of 1 μ at 1 cps and less than 0.1 μ at 10 cps.

Later tests at stations 2 and 3 were made with detectors having somewhat increased response to longer periods, and these showed displacements of up to a few microns at periods of 3 to 5 sec. These periods are in the range of storm microseisms, and it should not be surprising that relatively strong motion would be recorded. Unfortunately, however, the measurements are not wholly reliable because of electronic difficulties encountered at these stations. It appears now that in these two units the power transistors were not properly connected to the wall of the pressure case and as a result overheated the amplifier B batteries, causing some instability.

Station 2. During the monitoring period between seismic shots at this station, body waves from an earthquake at 69°N, 77.5°W, near the Panama-Colombia border, were recorded. The shock occurred on September 19, 1960, and was of magnitude 6 to 6¼ on the Richter scale.

P WAVES

19 SEPT., 1960

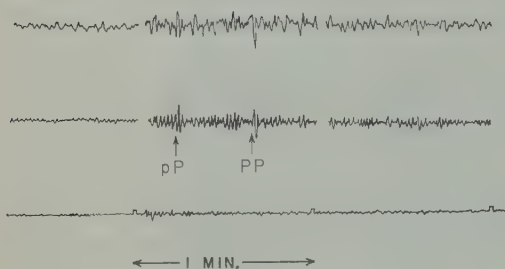


Fig. 13. Earthquake recorded by bottom seismograph (upper trace unfiltered; middle trace filtered 0.75 to 1.5 cps) and by Palisades short-period vertical instrument (lower trace).

Figure 13 shows the bottom seismograph record, unfiltered and filtered 0.75 to 1.5 cps, compared with the record from the Palisades 1-sec vertical instrument. (A standard Benioff short-period vertical seismograph, $T_0 = 1$ sec, $T_s = 0.2$ sec, both critically damped, maximum magnification 50,000 at 0.2 sec). The epicentral distance was 2800 km from the bottom seismograph and 3700 km from Palisades.

In recording the bottom seismograph data the minute marks were made by interrupting the signal for a few seconds, and unfortunately the beginning of the P wave occurred during one of these periods. The arrival shortly after P comes at about the proper time for pP or sP , and the arrival at about 40 seconds after P is apparently PP . The sharpness of this phase is noteworthy; it suggests that P would have been very clear and prominent had a more refined timing system been employed. Unfortunately, the end of the reel of magnetic tape arrived a few seconds before S .

Two refraction profiles were recorded at this station—one line shot north 52 km and the other shot south 268 km from the seismograph position. The structure determined by these profiles is of the type which has been shown to be typical of ocean basins by many investigators. The principal crustal layer (6.5 km/sec) and the mantle (8.0 km/sec) were clearly recorded. Exceptionally strong crustal shear waves were recorded, giving a velocity of 3.65 km/sec. A few arrivals which are probably mantle shear waves can be picked in the long profile. The indicated apparent velocity is 4.6 km/sec. The crustal velocities recorded here (both compressional and shear) are determined by end-to-

profile arrangement and are probably close to true velocities. Although the mantle velocities are from an unreversed profile, the values do not indicate any appreciable effect of slowness.

It is noteworthy that very strong crustal shear waves were recorded at this station whereas only a few rather weak mantle shear waves could be picked. At station 1 the mantle shear waves were much stronger than those from the crustal layer. This difference is apparently the effect of the character of the various interfaces. Ewing and Ewing [1959] noted that a marked difference in the strength of crustal shear waves existed between profiles which were quite similar in other respects and likewise attributed this to horizontal variations in the upper crustal interface.

Another observation of some significance in the long profile recorded here is the absence of any noticeable focusing or unusual diffusion of energy in the mantle P waves as the shot range was increased. The mantle arrivals were never exceptionally strong; and, from the range at which this phase became a first arrival (about 45 km) out to the maximum range, the amplitudes decreased at a relatively even rate.

Station 3. At this station the bottom seismograph was put down in the southern part of the Sigsbee deep, in the Gulf of Mexico. Shots were fired on azimuths of 090° and 300° from the receiving position. In the first of these profiles, ground-wave arrivals were measured at ranges out to 230 km (153 sec of water-wave time). In the second profile, with the same charge size being used, ground-wave arrivals could not be recorded beyond a range of 55 km. In this area, 45 to 50 sec is about the range within which mantle arrivals can be seen as first events on the seismograms, and the ray paths associated with these arrivals would include only the upper part of the mantle. Hence it would appear that these results indicate significant variations in the crust-mantle boundary over relatively short horizontal distances.

In the first profile at this station (090° azimuth) a pronounced focusing of energy was observed in the mantle arrivals at a range of 150 to 170 km. This can be seen in Figure 14 which shows records of mantle P waves from shots at ranges between 100 and 235 km. The shots were 120-kg depth charges, and all recorded

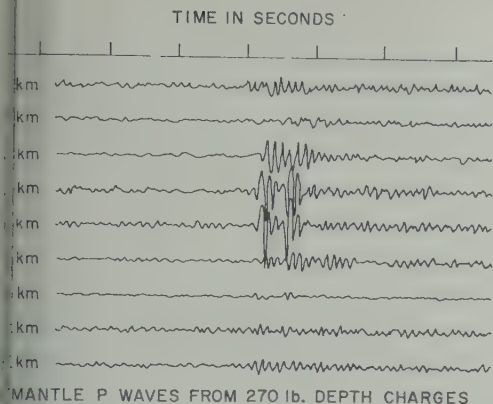


Fig. 14. Refracted mantle arrivals from profile 1 in the Gulf of Mexico showing focusing of energy at ranges of 150 to 170 km.

made with the same amplifier gain settings. Therefore there is no doubt that the focusing is a structural effect such as might result from a velocity gradient in the upper mantle or a focusing configuration in the crust-mantle interface.

From previous refraction studies in this area [Ewing, Antoine, and Ewing, 1960] it is known that the crustal thickness along this profile is appreciably greater than that along the second profile, which was shot northwestward across the main basin of the Gulf of Mexico. The first profile, in fact, entirely in the transitional region between oceanic and continental structure, whereas the major part of the second profile crosses structure that is more typical of ocean basins. Station 2, also in an oceanic area, did not give any evidence of focusing, so there is some reason to attach significance to these results, limited though they are. Because amplitudes as well as arrival times can be used, additional measurements of this type should give more definitive information about crustal and upper-mantle structure than has been available previously.

SUMMARY

The principal results derived from these tests are the following:

1. The system of placing a receiving installation on the sea floor and acoustically telemetering the information to a surface ship has proved satisfactory.
2. Body waves from an earthquake have been recorded, indicating that a good signal-to-noise ratio was achieved by receiving on the sea floor.

3. Useful data on background noise in the short-period range have been obtained.

4. Long refraction profiles have given additional measurements of crustal and mantle *P* and *S* waves.

5. Variation in the amount of energy propagated along various refraction paths indicates appreciable differences in interface characteristics from one area to another.

Acknowledgments. The work reported here was supported by the U. S. Navy under contracts NObsr 64547, Nonr-266(48), and Nonr-266(65), G. B. Tirey, J. F. Hennion, D. E. Koelsch, S. M. Gerber, A. C. Hubbard, J. L. Worzel, and D. D. Prentiss have made major contributions to many phases of the work. We wish particularly to acknowledge the important assistance rendered by C. Hartdegen and G. R. Hamilton of the Columbia University Geophysical Field Station in Bermuda, both in personal participation and in making available the services of R. V. *Sir Horace Lamb* and her crew. For the Gulf of Mexico operation, we wish to thank H. McLellan and J. W. Antoine for their support with R. V. *Hidalgo*.

Mr. Charles M. Grace's gift of the schooner *Grace* (formerly *Crishaza*) greatly facilitated the work, for it was aboard *Grace* that much of the testing of the instrument was done.

Also participating in one or more phases of this work were A. Stockel, R. Zaubere, and P. Kunsman of Lamont Observatory and F. S. Wendt of R. C. A. Princeton Laboratories.

B. Luskin played a leading part in the design and construction of the telemetering system and contributed effectively to the test at station 1.

REFERENCES

- Brune, J. N., and J. Oliver, The seismic noise of the earth's surface, *Bull. Seism. Soc. Am.*, **49**(4), 349-353, 1959.
- Dorman, J., M. Ewing, and J. Oliver, Study of shear-velocity distribution in the upper mantle by mantle Rayleigh waves, *Bull. Seism. Soc. Am.*, **50**(1), 87-115, 1960.
- Ewing, J., J. Antoine, and M. Ewing, Geophysical measurements in the western Caribbean Sea and in the Gulf of Mexico, *J. Geophys. Research*, **65**, 4087-4126, 1960.
- Ewing, J., and M. Ewing, Seismic-refraction measurements in the Atlantic Ocean basins, in the Mediterranean Sea, on the mid-Atlantic ridge, and in the Norwegian Sea, *Bull. Geol. Soc. Am.*, **70**, 291-317, 1959.
- Ewing, M., and F. Press, Propagation of elastic waves in the ocean with reference to microseisms, *Pontificiae Academiae Scientiarum Scripta Varia*, **12**, 121-127, 1952.
- Ewing, M., and A. C. Vine, Deep-sea measurements without wires or cables, *Trans. Am. Geophys. Union*, part 1, 248-251, 1938.

- Ewing, M. C. F., Woodlark, A. C., Hise, and J. L. Wengel, Recent results in volcanologic geophysics. *Soil. Geol. Soc. Am.*, 57, 400-406, 1963.
- Hise, J. B., C. B. Joffe, H. E. Johnson, and S. Bergstrom, Seismic refraction observations north of the Hawaiian arch. *Soil. Geol. Soc. Am.*, 47, 291-306, 1952.
- Kane, S. and M. Ewing, Seismicity patterns in the Atlantic Ocean. 7. 30 Ocean basins west of Bermuda. *Soil. Geol. Soc.*, 47, 473-500, 1953.
- Manuscript received July 21, 1963 (revised August 26, 1964.)

An Evaluation of a Signal-Summing Technique for Improving the Signal-to-Noise Ratios for Seismic Events

L. H. KOOPMANS

Sandia Corporation
Albuquerque, New Mexico

Abstract. Formulas are derived for evaluating as a function of frequency the improvement in the root mean square signal-to-noise ratio for seismic events obtainable by the technique of summing the trace amplitudes recorded at a number of closely spaced seismometer stations. A modification of this technique whereby the signals are corrected for phase shift before summing is also studied. The theory is applied to noise data from a three-station network for the ranges of periods 6.7 to 20 sec and 0.25 to 1 sec. The signals in the short-period range are taken to be *P* and *S* body waves. Rayleigh waves are used in the long-period range. Improvement comparable to and occasionally in excess of that expected for uncorrelated noise is indicated for the modified technique in the short-period range, whereas little improvement is obtained in the long-period range for the given station spacing and noise conditions.

1. INTRODUCTION

The merging of techniques of time series analysis from the field of mathematical statistics and engineering with problems in geophysics has provided both a rich field of applications for time series theory and a valuable arsenal of analytic tools for the geophysicist. One of the most important among these tools is power spectrum analysis—the harmonic analysis of time functions which, when considered over all time, have infinite energy but finite power. (In this case power has the dimension of squared amplitude.) For deterministic functions the first comprehensive treatment of power spectrum analysis was given by Wiener [1930]. When the function is considered a sample function of a stationary stochastic process an analogous harmonic analysis is available because of the work of many authors [see bibliography in Bendat and Rosenblatt, 1957]. However, this work was not generally available to the geophysicist until Tukey's [1949] paper was published, in which the sampling theory for random stochastic processes was covered in detail and criteria for designing and interpreting power spectrum analyses were given. This theory was enlarged upon by Blackman and Tukey [1958] and extended to pairs of gaussian processes by Goodman [1957], who derived approximations to the statistical distributions of many functions relating the two processes.

Power spectrum analysis was not brought to geophysics as a routine analytic tool until high-

speed digital computers came into use. In recent years machine programs for performing spectrum analyses on single or bivariate time series have become available for such machines as the IBM 704 [e.g. Clark, 1958] and data processing packages have been developed in conjunction with these programs for performing various kinds of filtering as well as such routine data-handling tasks as the graphing of spectra on a standard output printer [e.g. Young, 1961].

The present paper was motivated by the desire to apply power spectrum techniques to the problem of monitoring seismic signals for the detection of underground nuclear explosions. Kogan, Pasechnik, and Sultanov [1959] and others have studied the problem of distinguishing between earthquakes and nuclear explosions on the basis of the distribution of signal power with frequency. The indications are that power spectrum analysis, although not providing a complete solution to the problem, will nevertheless be a valuable auxiliary tool for use in the routine analysis of 'suspect' signals.

Of great economic and political concern is the spacing and hence the number of manned monitoring stations required to cover a given land area. This in turn depends on the relationship of expected signal strength to the intrinsic background noise of the area. For power spectrum analysis the natural measure of this relationship is the ratio of root mean square (rms) signal amplitude to rms noise amplitude

computed for some interval of time containing the signal or possibly some particular portion of it, e.g. the portion containing only *P* waves. Any method of increasing the signal-to-noise ratio and thus decreasing the required number of stations would be of considerable importance. The purpose of this paper is to derive analytic expressions for the evaluation of one such technique when records of noise data are available for the region of interest.

The method is as follows: Each monitor station would consist of one manned location and $N-1$ unmanned satellite sites consisting of seismometers and amplifying equipment sufficient to receive and transmit signals in the frequency ranges of interest. (The selection of the number N will not be discussed here, as it depends, in part, on economic considerations outside the scope of the present work.) The signals from the N locations would be simultaneously fed into an electronic device at the manned site which would add together the instantaneous signal amplitudes and record the result for analysis. The N sites would be located sufficiently close together so that very little cancellation of the desired signal due to phase shifting and attenuation would take place when the traces were added. The rms amplitude of the desired signal in this case should be roughly N times as large as that for a single station. On the other hand, if the background noise at each site were uncorrelated with that at each of the other sites, when the traces were added random cancellations would cause the rms noise amplitude to increase on the average to only \sqrt{N} times the value at a single site. This would provide an improvement factor of \sqrt{N} in the signal-to-noise ratio.

Of course, the ideal situation of uncorrelated noises is seldom achieved in nature, and the degree to which it is achieved depends strongly on frequency. Moreover, an attempt to decrease the noise correlation by varying the separation between satellite stations is counterbalanced to a degree by the increased interference of the desired signals themselves.

In this paper formulas for the evaluation of the signal-summing technique are derived for fixed but arbitrary station configurations for which noise records are assumed available. No attempt is made to propose optimal spacings or configurations for the satellite stations, as this

would depend upon a knowledge of the variation of noise correlation with distance not presently available.

The present technique for the improvement in signal-to-noise ratio is not new. A descriptive discussion of a similar method under the name of mixing was given by *Frank and Doty* [1951] for station spacings and frequency ranges pertinent to seismic prospecting. They were able to arrive at rough estimates of the achievable improvement by carrying out an experiment in which weighted sums of observed noise traces with artificially superimposed signals were constructed. However this procedure was empirical and little more could be concluded than that, in the frequency range they considered, the method seemed to have real merit.

Holtzmann [1960] attempted an analytic evaluation of the technique through a harmonic analysis of a nondispersive signal plus noise in spatial frequency components. By representing the mixing as a convolution in the space variable a simple harmonic representation is maintained. However, the functional form assumed for the noise produces perfect noise correlation between all pairs of stations regardless of their spatial separation. This situation is not borne out by experiment nor is the mathematically inconsistent conclusion that the correlation drops to zero as the station separation becomes infinite. It is felt by this author that *Holtzmann's* treatment is overly restricted by simplifying assumptions and is inadequate for a realistic evaluation of the summing procedure. Assumptions concerning the degree of noise correlation are avoided in this paper.

The remainder of the paper is arranged as follows: Section 2 contains the definitions and assumptions that are used throughout. In section 3 formulas are derived for assessing the achievable improvement in signal-to-noise ratio using the present scheme and taking into account the magnitudes of the noise correlation as functions of frequency and the interference of the desired signals when dispersion but attenuation may take place between stations. In section 4 an evaluation of the improvement in signal-to-noise ratio is given for a modification of this scheme wherein each signal is corrected for phase shift before the summing operation. In section 5 the formulas of sections 3 and 4 are applied to data from a trio of seismometer

near Albuquerque, New Mexico. The accuracy of these stations is within the range of the capability of the theory, making it possible to estimate the improvement in signal-to-noise ratio for three station separations.

2. DEFINITIONS AND ASSUMPTIONS

Let the N seismometer locations be labeled in a convenient fashion from 1 to N and let $S_\mu(t)$ be the output of a given type of seismometer system (e.g. high-frequency system) at time t for the μ th station. If the time origin is taken near the arrival time of a given 'suspect' disturbance at the stations in the array, the total $X_\mu(t)$ will have the form

$$X_\mu(t) = S_\mu(t) + N_\mu(t) \quad \mu = 1, 2, \dots, N$$

where $S_\mu(t)$ is the component of the output attributable to the disturbance (previously referred to as the desired signal) and $N_\mu(t)$ is an undesired but ever-present background noise. For convenience it will be assumed that the signals are all observed over the fixed time interval $0 \leq t \leq A$. The summing operation produces the new trace

$$X(t) = \sum_{\mu=1}^N X_\mu(t)$$

the resultant decomposition $X(t) = S(t) + N(t)$ into desired signal and noise. The natural model for the $S_\mu(t)$ component is that of a deterministic, transient time function—a function which when considered over all time $-\infty < t < \infty$ would have finite energy but infinite power. That is,

$$\int_{-\infty}^{\infty} S_\mu^2(t) dt < \infty$$

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T S_\mu^2(t) dt = 0$$

On the other hand, to characterize the random nature of seismic noise, it is natural to assume that the $N_\mu(t)$ component is from a stationary time series which implies that¹

The integrals and limits for the stochastic processes will be taken in the mean square sense which requires stationarity of the processes up to fourth moments. Conditions under which the elements to be made are valid may be found in standard probability texts such as those of Grenander and Rosenblatt [1957] or Doob [1953].

$$\int_{-\infty}^{\infty} N_\mu^2(t) dt = \infty$$

but

$$0 < \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T N_\mu^2(t) dt < \infty$$

It is well known that the usual forms of harmonic analysis for transient and nontransient phenomena such as these are incompatible over the infinite time axis. This dilemma is only apparent, however, since we are always restricted to a finite interval of time for which both transient and nontransient phenomena have finite energy and nonzero power. When power spectrum analysis is applied to a transient signal over a finite interval, it produces a smoothed version of the squared modulus of its Fourier transform where the signal is taken to be zero outside of the interval. For the desired signal this is indeed the quantity of interest. However, since the noise is random in character, only average properties (in the ensemble sense) have relevance to an evaluation of the noise level of a region. Under weak restrictions, ensemble averages for stationary time series may be replaced by time averages over the infinite time axis. A time average over $0 \leq t \leq A$ in this case provides a statistical estimate of the corresponding ensemble average.

To avoid introducing unnecessary notation, the theory will be formulated in terms of time averages over the infinite time axis. This is accomplished by replacing the transient signals $S_\mu(t)$ by nontransient ones having positive power as well as the same average properties over all time as the original signals had over $0 \leq t \leq A$.

As a matter of convenience it is assumed that all instruments are balanced so that no d-c component is present in either signal or noise. That is,

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T S_\mu(t) dt = 0$$

and

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T N_\mu(t) dt = 0$$

for $\mu = 1, 2, \dots, N$. Let

$$\varphi_{S_\mu S_\nu}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T S_\mu(t + \tau) S_\nu(t) dt$$

$$\varphi_{N_\mu N_\nu}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T N_\mu(t + \tau) N_\nu(t) dt$$

and

$$\varphi_{S_\mu N_\nu}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T S_\mu(t + \tau) N_\nu(t) dt$$

for $-\infty < \tau < \infty$ and $\mu = 1, 2, \dots, N$. These are the cross covariance functions of signal and noise and together contain all the power information relative to the N seismic stations. The cross covariance of a signal with itself is called the autocovariance function, and the notation

$$\varphi_{S_\mu S_\mu}(\tau) = \varphi_{S_\mu}(\tau) \quad \text{and} \quad \varphi_{N_\mu N_\mu}(\tau) = \varphi_{N_\mu}(\tau)$$

is adopted. It is assumed that the spectra of the signals and noises are absolutely continuous. Then, in the frequency domain, the power information is contained in the cross spectral densities which are the Fourier transforms of the cross covariance functions:

$$f_{S_\mu S_\nu}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} \varphi_{S_\mu S_\nu}(\tau) d\tau$$

and

$$f_{N_\mu N_\nu}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} \varphi_{N_\mu N_\nu}(\tau) d\tau$$

for $-\infty < \omega < \infty$. The Fourier transform of the autocovariance function of a signal is called simply the spectral density of the signal, and the notation

$$f_{S_\mu S_\mu}(\omega) = f_{S_\mu}(\omega) \quad \text{and} \quad f_{N_\mu N_\mu}(\omega) = f_{N_\mu}(\omega)$$

is used. It is easily verified that these functions are real values.

The assumption that the spectra are absolutely continuous excludes the possibility of spectral lines (frequencies of positive power). It can be shown that this assumption also guarantees that the signals are uncorrelated with the noises. That is, $\varphi_{S_\mu N_\nu}(\tau) = 0$ for all $\mu, \nu = 1, 2, \dots, N$. This in turn implies that $S(t)$ and $N(t)$ are uncorrelated.

The spectral densities contain the power information for the system of N stations in a slightly more intuitive form than do the cross covariance functions. For example, in the signal $S_\mu(t)$ the power that is attributable to its harmonic components with frequencies between ω_0 and ω_1 is

$$\int_{\omega_0}^{\omega_1} f_{S_\mu}(\omega) d\omega$$

The power and phase information for, $S_\mu(t)$ and $S_\nu(t)$ are contained in $f_{S_\mu S_\nu}(\omega)$ in somewhat more disguised form. Insight into this will be obtained in the course of the derivations of sections 3 and 4.

It is assumed that the earth in the region containing the N stations together with the instrument systems behaves like a linear filter. Thus the signals $S_\mu(t)$ are related to $S_1(t)$ by equations

$$S_\mu(t) = \int_{-\infty}^{\infty} h_\mu(u) S_1(t - u) du$$

$$\mu = 2, 3, \dots, N$$

The function $h_\mu(u)$ is the impulse response of the earth between the first and the μ th station and the instrument systems in these stations. Generalized functions will be used in the succeeding discussions in order to represent idealized filter characteristics such as no dispersion and no signal attenuation. Thus, for example, a linear filter which 'does nothing' to $S_1(t)$ may be represented by the impulse response $\delta(u)$, the Dirac delta function, since

$$S_1(t) = \int_{-\infty}^{\infty} \delta(u) S_1(t - u) du$$

Equation 1 may then be extended to include index $\mu = 1$ by setting $h_1(u) = \delta(u)$.

The Fourier transform $H_\mu(\omega)$ of $h_\mu(u)$ is the transfer function of the μ th linear filter, and characterizes the properties of the filter as a function of frequency.

$H_\mu(\omega)$ is, in general, a complex valued function which may be written in the form $H_\mu(\omega) = |H_\mu(\omega)| e^{-i\theta_\mu(\omega)}$. The two factors on the right-hand side of this equation contain, respectively, the attenuation and phase shift of $S_\mu(t)$ relative to $S_1(t)$. For example, the transfer function of the 'do nothing' filter is $H_1(\omega) \equiv 1$, which corresponds to no attenuation ($|H_1(\omega)| = 1$) and no phase shift ($\theta_1(\omega) = 0$) for any frequency. In general, the condition of no attenuation in signal strength corresponds to

$$|H_\mu(\omega)| = 1$$

for $-\infty < \omega < \infty$. For the frequency range and station spacing contemplated in this study, this will be valid to a good degree of approximation for all stations and instrument systems. Thus (2) will be assumed to hold hereafter.

will also be assumed that the phase characteristics of all instrument systems are the same. Then, since the transformation from $S_1(t)$ to $S_\mu(t)$ is from the output of the first instrument to the output of the μ th, the instrument phase shifts will cancel and $\theta_\mu(\omega)$ will represent the phase characteristics of the earth alone.

For a nondispersive signal, the phase angles $\theta_\mu(\omega)$ have particularly simple forms. Suppose $S_\mu(t)$ is obtained from $S_1(t)$ by a simple time shift;

$$S_\mu(t) = S_1(t - t_{1\mu})$$

where $t_{1\mu}$ is the travel time of the signal between stations 1 and μ . This represents a linear transformation of the form given in (1) with $h_\mu(u) = \delta(u - t_{1\mu})$. Again δ is the Dirac delta function.

$$\begin{aligned} S_\mu(\omega) &= \int_{-\infty}^{\infty} e^{-i\omega u} h_\mu(u) du \\ &= \int_{-\infty}^{\infty} e^{-i\omega u} \delta(u - t_{1\mu}) du \end{aligned}$$

where $v = u - t_{1\mu}$. Then

$$S_\mu(\omega) = e^{-i\omega t_{1\mu}} \int_{-\infty}^{\infty} e^{-i\omega v} \delta(v) dv = e^{-i\omega t_{1\mu}}$$

It follows from this that

$$\theta_\mu(\omega) = t_{1\mu}\omega$$

If V is the (constant) phase velocity, then $t_{1\mu} = d_{1\mu}/V$. Hence

$$\theta_\mu(\omega) = (d_{1\mu}/V)\omega \quad (3)$$

In the case of a plane wave, $d_{1\mu}$ is the distance between the projections of the first and μ th stations on a line perpendicular to the wave front. For a dispersive plane wave the velocity depends on frequency, so that

$$\theta_\mu(\omega) = \frac{d_{1\mu}}{V(\omega)} \omega \quad (4)$$

Although other spatial coordinate systems and wave profiles are not excluded by the theory, only cartesian coordinates and plane waves will be considered in this paper.

The theory will be developed for the special case in which the noise spectral densities are equal:

$$f_{N_\mu}(\omega) = f_{N_1}(\omega) \quad (5)$$

for $\mu = 2, 3, \dots, N$ and all ω . This assumption implies no restriction on the noise correlations between stations and will hold to a good degree of approximation for many areas, frequency ranges, and station separations. In particular, it is a good approximation in the present case. When the assumption is badly violated, the ensuing derivations can be generalized without difficulty but will lead to more complicated expressions.

If $f_S(\omega)$ and $f_N(\omega)$ are the spectral densities of the summed signals and noises, the resultant signal-to-noise ratio for the frequency interval $\omega_0 \leq \omega \leq \omega_1$ is

$$\left[\frac{\int_{\omega_0}^{\omega_1} f_S(\lambda) d\lambda}{\int_{\omega_0}^{\omega_1} f_N(\lambda) d\lambda} \right]^{1/2}$$

This is to be compared with the original signal-to-noise ratios

$$\left[\frac{\int_{\omega_0}^{\omega_1} f_{S_\mu}(\lambda) d\lambda}{\int_{\omega_0}^{\omega_1} f_{N_\mu}(\lambda) d\lambda} \right]^{1/2}$$

where $\mu = 1, 2, \dots, N$. As will be shown later, $f_{S_\mu}(\omega) = f_{S_1}(\omega)$ for $\mu = 2, 3, \dots, N$. Hence, the signal-to-noise ratios for all stations are equal, and it suffices to compare the ratio for the sum with that for the first station.

A considerable reduction in the complexity of the formulas to be derived in the following sections will be achieved by selecting the interval widths $\Delta\omega = \omega_1 - \omega_0$ so small that the cross spectral densities of both signals and noise are effectively constant over each interval. Then the integrals of these densities will be closely approximated by expressions of the form

$$\int_{\omega_0}^{\omega_1} f_{S_\mu S_\nu}(\lambda) d\lambda \cong f_{S_\mu S_\nu}(\omega) \Delta\omega,$$

etc., where $\omega = (\omega_0 + \omega_1)/2$.

The improvement factor relating the signal-to-noise ratio of the summed signals to that of the originals is then defined to be

$$K(\omega) = \left[\frac{f_S(\omega)}{f_{N_1}(\omega)} \right]^{1/2} \quad (6)$$

It is to be emphasized that $K(\omega)$ does not pertain to the single frequency ω but rather

represents an approximation to the exact expression involving integrals over the interval of length $\Delta\omega$ centered at ω .

3. CALCULATION OF THE IMPROVEMENT FACTOR OF THE SIGNAL-TO-NOISE RATIO

In this section it is shown that

$$K(\omega) = \left[\frac{N + 2 \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N \cos \Delta_{\mu\nu}(\omega)}{N + 2 \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N c_{N\mu N\nu}(\omega)} \right]^{1/2} \quad (7)$$

where $\Delta_{\mu\nu}(\omega) = \theta_{\mu}(\omega) - \theta_{\nu}(\omega)$ and $c_{N\mu N\nu}(\omega) = R [f_{N\mu N\nu}(\omega)]/f_{N_1}(\omega)$. ($R []$ denotes the real part of the quantity in brackets.)

To verify (7) the spectral density function of the summed signal $S(t)$ is first calculated:

$$S(t) = \sum_{\mu=1}^N S_{\mu}(t) = \sum_{\mu=1}^N \int_{-\infty}^{\infty} h_{\mu}(u) S_1(t-u) du$$

Thus

$$\begin{aligned} \frac{1}{2T} \int_{-T}^T S(t+\tau) S(t) dt \\ = \frac{1}{2T} \int_{-T}^T \sum_{\mu=1}^N \sum_{\nu=1}^N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{\mu}(u) h_{\nu}(v) \\ \cdot S_1(t+\tau-u) S_1(t-v) du dv dt \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{\mu=1}^N \sum_{\nu=1}^N h_{\mu}(u) h_{\nu}(v) \\ \cdot \left[\frac{1}{2T} \int_{-T}^T S_1(t+\tau-u) S_1(t-v) dt \right] du dv \end{aligned}$$

Passing to the limit as $T \rightarrow \infty$, this expression yields

$$\varphi_S(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{\mu=1}^N \sum_{\nu=1}^N h_{\mu}(u) h_{\nu}(v) \cdot \varphi_{S_1}(\tau-u+v) du dv$$

Now, taking the Fourier transform of both sides of this equation, the spectral density of $S(t)$ is obtained:

$$\begin{aligned} f_S(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{\mu=1}^N \sum_{\nu=1}^N h_{\mu}(u) h_{\nu}(v) \\ \cdot \left\{ \int_{-\infty}^{\infty} e^{-i\omega\tau} \varphi_{S_1}(\tau-u+v) d\tau \right\} du dv \end{aligned}$$

By the change of variable $\sigma = \tau - u + v$, the quantity in braces becomes

$$\int_{-\infty}^{\infty} e^{-i\omega(\sigma+u-v)} \varphi_{S_1}(\sigma) d\sigma = e^{-i\omega(u-v)} f_{S_1}(\omega)$$

Thus,

$$\begin{aligned} f_S(\omega) &= f_{S_1}(\omega) \sum_{\mu=1}^N \sum_{\nu=1}^N \int_{-\infty}^{\infty} e^{-i\omega u} h_{\mu}(u) du \\ &\cdot \int_{-\infty}^{\infty} e^{i\omega v} h_{\nu}(v) dv \\ &= f_{S_1}(\omega) \sum_{\mu=1}^N \sum_{\nu=1}^N H_{\mu}(\omega) \overline{H_{\nu}(\omega)} \end{aligned}$$

where the bar denotes complex conjugate. Hence, by assumption, $H_{\mu}(\omega) = e^{-i\theta_{\mu}(\omega)}$. Hence,

$$\begin{aligned} \sum_{\mu=1}^N \sum_{\nu=1}^N H_{\mu}(\omega) \overline{H_{\nu}(\omega)} &= \sum_{\mu=1}^N |H_{\mu}(\omega)|^2 \\ &+ \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N [H_{\mu}(\omega) \overline{H_{\nu}(\omega)} + \overline{H_{\mu}(\omega)} H_{\nu}(\omega)] \\ &= N + 2 \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N \cos \Delta_{\mu\nu}(\omega) \end{aligned}$$

This quantity is the ratio of $f_S(\omega)$ to $f_{S_1}(\omega)$ and because of (6) it appears as the numerator of the expression in (7). To obtain the denominator of this expression it is necessary to calculate the spectral density of

$$\begin{aligned} N(t) &= \sum_{\mu=1}^N N_{\mu}(t) \\ \frac{1}{2T} \int_{-T}^T N(t+\tau) N(t) dt \\ &= \sum_{\mu=1}^N \sum_{\nu=1}^N \frac{1}{2T} \int_{-T}^T N_{\mu}(t+\tau) N_{\nu}(t) dt \end{aligned}$$

Thus, letting $T \rightarrow \infty$ yields

$$\varphi_N(\tau) = \sum_{\mu=1}^N \sum_{\nu=1}^N \varphi_{N_{\mu} N_{\nu}}(\tau)$$

and taking Fourier transforms yields

$$f_N(\omega) = \sum_{\mu=1}^N \sum_{\nu=1}^N f_{N_{\mu} N_{\nu}}(\omega)$$

But

$$\begin{aligned} \varphi_{N_{\mu} N_{\nu}}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T N_{\mu}(t+\tau) N_{\nu}(t) dt \\ &= \varphi_{N_{\mu} N_{\nu}}(\tau) \end{aligned}$$

Thus

$$\begin{aligned} \varphi_{N\mu N\nu}(\omega) &= \int_{-\infty}^{\infty} e^{-i\omega\tau} \varphi_{N\mu N\nu}(-\tau) d\tau \\ &= \int_{-\infty}^{\infty} e^{i\omega\tau} \varphi_{N\mu N\nu}(\tau) d\tau = \overline{f_{N\mu N\nu}(\omega)} \end{aligned}$$

or, rewriting (9) in the form

$$\begin{aligned} \varphi_{N\mu N\nu}(\omega) &= \sum_{\mu=1}^N f_{N\mu}(\omega) \\ &+ \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N [f_{N\mu N\nu}(\omega) + f_{N\nu N\mu}(\omega)] \end{aligned}$$

and the spectral density of $N(t)$:

$$f_N(\omega) = f_{N1}(\omega) \left[N + 2 \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N c_{N\mu N\nu}(\omega) \right]$$

where $c_{N\mu N\nu}(\omega)$ is defined by

$$c_{N\mu N\nu}(\omega) = f_{N1}(\omega) [c_{N\mu N\nu}(\omega) + iq_{N\mu N\nu}(\omega)] \quad (10)$$

which completes the verification of (7).

The quantities $c_{N\mu N\nu}(\omega)$ and $q_{N\mu N\nu}(\omega)$ are normalized versions of the co-spectral density and quadrature spectral density defined by Goodman [1957]. The actual normalizing factor is $[f_N(\omega)f_{N1}(\omega)]^{-1/2}$ which, because of equation 5, reduces to $f_{N1}(\omega)^{-1}$. Estimates of these quantities are included in the outputs of all cross-spectrum analysis programs; consequently, little additional computation beyond the power spectrum is involved in applying (7) to an actual station configuration. In the calculations for section 5 we used the actual rather than the assumed normalizing factor in the computations of $c_{N\mu N\nu}(\omega)$ and $q_{N\mu N\nu}(\omega)$.

Equation 7 provides an estimate of the improvement in signal-to-noise ratio for a given signal direction. In general, to assess the overall performance of the signal-summing technique for a particular station configuration, one is interested in a criterion which is independent of signal direction. Many such criteria are possible, and no natural ones will now be suggested.

A. If one is interested in the improvement obtainable in the worst (best) possible case that could occur, the signal direction is selected to minimize (maximize) $K(\omega)$ as a function of signal direction. Let (x_μ, y_μ) be the coordinates of the μ th station in a conveniently selected Cartesian coordinate system and let the first station be located at the origin. Then if the normal vector in the direction of propagation of

a plane wave makes an angle α with the x axis, the projected distance $d_{1\mu}$ of equations 3 and 4 is

$$d_{1\mu}(\alpha) = x_\mu \cos \alpha + y_\mu \sin \alpha$$

This evaluation of distance and equation 7 yield

$$\begin{aligned} \Delta_{\mu\nu}(\omega, \alpha) &= \frac{\omega}{V(\omega)} \\ &\cdot [(x_\mu - x_\nu) \cos \alpha + (y_\mu - y_\nu) \sin \alpha] \quad (11) \end{aligned}$$

If $K(\omega, \alpha)$ denotes the improvement ratio with (11) replacing $\Delta_{\mu\nu}(\omega)$ in (7), then $K(\omega, \alpha)$ is clearly maximized or minimized for those values of α for which the function

$$G(\omega, \alpha) = \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N \cos \Delta_{\mu\nu}(\omega, \alpha) \quad (12)$$

is maximized or minimized. If we set the partial derivative of G with respect to α equal to zero and use the small angle approximation, $\sin \theta \approx \theta$, the approximate values of α for which the maxima and minima of $K(\omega, \alpha)$ occur are the solutions of the equation

$$\tan 2\alpha = \frac{2 \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N (x_\mu - x_\nu)(y_\mu - y_\nu)}{\sum_{\mu=1}^N \sum_{\nu=\mu+1}^N (x_\mu - x_\nu)^2 - (y_\mu - y_\nu)^2} \quad (13)$$

For stations located along a straight line we may choose the coordinate axes so that $y_\mu = 0$, $\mu = 1, 2, \dots, N$. The signal directions which minimize or maximize $K(\omega, \alpha)$ are then included among the angles satisfying the equation $\tan 2\alpha = 0$, i.e. $\alpha = 0, \pm\pi/2, \pm\pi, \dots$. It is seen that $\alpha = 0, \pm\pi$ are minima, and $\alpha = \pm\pi/2$ are maxima since

$$\begin{aligned} G(\omega, 0) &= G(\omega, \pm\pi) \\ &= \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N \cos \frac{\omega}{V(\omega)} (x_\mu - x_\nu) \\ &\leq G(\omega, \pm\pi/2) = \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N 1 = \frac{N(N-1)}{2} \end{aligned}$$

This result coincides with intuition which tells us that it is most favorable for the signal to be received simultaneously at all N stations and least favorable for the wave front to be perpendicular to the line of stations. Note that (13) does not involve ω ; thus the optimal signal

directions, once obtained, will be the same for all frequencies.

If the interesting signals were restricted to some particular range of directions, the maxima and minima of (12) would be computed subject to the corresponding restrictions on the variable α . Only minor alterations of the above discussion must be made in this case; hence the topic will not be pursued further.

B. If the relative frequencies $\Pi(\alpha)$ of observed signal directions are known, a more realistic evaluation of the improvement ratio might be the expected or average value

$$K^*(\omega) = \int_0^{2\pi} K(\omega, \alpha) \Pi(\alpha) d\alpha$$

where $\Pi(\alpha')d\alpha$ is the probability or relative frequency with which signals are observed from directions $\alpha' \leq \alpha \leq \alpha' + d\alpha$. Thus $\Pi(\alpha) \geq 0$ and $\int_0^{2\pi} \Pi(\alpha) d\alpha = 1$. If all directions are equally probable, $\Pi(\alpha) = 1/2\pi$ and

$$K^*(\omega) = \frac{1}{2\pi} \int_0^{2\pi} K(\omega, \alpha) d\alpha$$

Since $\Pi(\alpha)$ is seldom known, the maximum and minimum values of $K(\omega)$ discussed in part A will be used in section 5. These values provide upper and lower bounds on $K^*(\omega)$ for any choice of $\Pi(\alpha)$ and, in fact, represent the most pessimistic and most optimistic choices of direction distributions;

$$\Pi_{\min}(\alpha) = \delta(\alpha - \alpha_{\min})$$

and

$$\Pi_{\max}(\alpha) = \delta(\alpha - \alpha_{\max})$$

where α_{\min} and α_{\max} are minima and maxima of $K(\omega, \alpha)$. These distributions concentrate all the probability at the worst possible and best possible directions, respectively.

4. EVALUATION OF A MODIFIED SIGNAL-SUMMING TECHNIQUE

A natural modification of the signal-summing techniques follows: As the signals are received at the monitoring station, they are recorded separately instead of being fed into the summing circuit. Then, from knowledge of the phase velocity distribution of the region, signal direction, and station configuration, linear transformations are performed on the signals $X_\mu(t)$, $\mu = 2, 3, \dots, N$, to correct the $S_\mu(t)$ com-

ponents back to $S_1(t)$. Finally, the traces are summed and the sum is recorded for further analysis. This procedure has the advantage of making the desired signals perfectly in phase at the station (to within the accuracy of the correction), but it has the disadvantage of requiring a considerable amount of additional data processing. Moreover, since noise in the low-frequency range is often attributable to coherent sources such as distant storm centers [see, for example, Gutenberg, 1958], by such correction the noise correlation between stations may be increased as well. The example in section 5 will illustrate this.

Let $w_\mu(u)$ be the impulse response function of the μ th correction filter, and let

$$Y_\mu(t) = \int_{-\infty}^{\infty} w_\mu(u) X_\mu(t - u) du$$

$$\mu = 1, 2, \dots$$

By equation 1,

$$X_\mu(t) = \int_{-\infty}^{\infty} h_\mu(u) S_1(t - u) du + N_\mu(t)$$

Thus, to correct the signals $S_\mu(t)$ back to $S_1(t)$ it is necessary to have

$$\begin{aligned} S_1(t) &= \int_{-\infty}^{\infty} w_\mu(v) S_\mu(t - v) dv \\ &= \int_{-\infty}^{\infty} w_\mu(v) dv \int_{-\infty}^{\infty} h_\mu(u) S_1(t - u) du \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} w_\mu(v) h_\mu(r - v) dv \right] S_1(t - r) dr \end{aligned}$$

This implies

$$\int_{-\infty}^{\infty} w_\mu(v) h_\mu(r - v) dv = \delta(r)$$

Hence, taking Fourier transforms on both sides of this equation, the transfer function of the μ th filter must satisfy the relation

$$W_\mu(\omega) H_\mu(\omega) \equiv 1$$

Since $H_\mu(\omega) = e^{-i\theta_\mu(\omega)}$, the correction transfer function is

$$W_\mu(\omega) = e^{i\theta_\mu(\omega)}$$

Then if perfect compensation is assumed, corrected traces are

$$= S_1(t) + \int_{-\infty}^{\infty} w_{\mu}(u) N_{\mu}(t-u) du$$

$$\mu = 1, 2, \dots, N.$$

will now be shown that the improvement for the modified summing technique is

$$K_M(\omega) = \frac{N}{\left[N + 2 \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N \{ \cos \Delta_{\mu\nu}(\omega) c_{N_{\mu}N_{\nu}}(\omega) - \sin \Delta_{\mu\nu}(\omega) q_{N_{\mu}N_{\nu}}(\omega) \} \right]^{1/2}} \quad (15)$$

before, $\Delta_{\mu\nu}(\omega) = \theta_{\mu}(\omega) - \theta_{\nu}(\omega)$, and $c_{N_{\mu}N_{\nu}}(\omega)$ and $q_{N_{\mu}N_{\nu}}(\omega)$ are as defined in (10). Let $S^*(t)$ and $N^*(t)$ be the signal and noise components of $Y(t) = \sum_{\mu=1}^N Y_{\mu}(t)$. Then, from equation 14, $S^*(t) = NS_1(t)$, and it is an easy matter to verify that

$$f_{S^*}(\omega) = N^2 f_{S_1}(\omega) \quad (16)$$

Moreover, by a computation similar to the leading to (8),

$$f_{N^*}(\omega) = \sum_{\mu=1}^N \sum_{\nu=1}^N W_{\mu}(\omega) \overline{W_{\nu}(\omega)} f_{N_{\mu}N_{\nu}}(\omega) \quad (17)$$

$$\begin{aligned} & \sum_{\nu=1}^N W_{\mu}(\omega) \overline{W_{\nu}(\omega)} f_{N_{\mu}N_{\nu}}(\omega) \\ &= \sum_{\mu=1}^N |W_{\mu}(\omega)|^2 f_{N_{\mu}}(\omega) \\ &+ \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N [W_{\mu}(\omega) \overline{W_{\nu}(\omega)} f_{N_{\mu}N_{\nu}}(\omega) \\ &+ \overline{W_{\nu}(\omega)} W_{\mu}(\omega) f_{N_{\nu}N_{\mu}}(\omega)] \end{aligned}$$

$$\begin{aligned} & f_{N_1}(\omega) \\ &+ 2 \sum_{\mu=1}^N \sum_{\nu=\mu+1}^N R[W_{\mu}(\omega) \overline{W_{\nu}(\omega)} f_{N_{\mu}N_{\nu}}(\omega) \\ &+ \overline{W_{\nu}(\omega)} W_{\mu}(\omega) f_{N_{\nu}N_{\mu}}(\omega)] \end{aligned}$$

$$\begin{aligned} &= f_{N_1}(\omega) [\cos \Delta_{\mu\nu}(\omega) + i \sin \Delta_{\mu\nu}(\omega)] \\ &\cdot [c_{N_{\mu}N_{\nu}}(\omega) + i q_{N_{\mu}N_{\nu}}(\omega)] \end{aligned}$$

$$\begin{aligned} & f_{N_1}(\omega) \overline{W_{\nu}(\omega)} f_{N_{\mu}N_{\nu}}(\omega) = f_{N_1}(\omega) \\ & \{ \cos \Delta_{\mu\nu}(\omega) c_{N_{\mu}N_{\nu}}(\omega) - \sin \Delta_{\mu\nu}(\omega) q_{N_{\mu}N_{\nu}}(\omega) \} \end{aligned}$$

Equation 14 is now obtained by applying definition 6 to equations 16 and 18.

Several methods for correcting dispersive waves for the properties of the earth and instrument response have been proposed in the literature.

One such method based on the cross correlation of an observed record with a function constructed from the phase velocity distribution is given by Aki [1960] along with an extensive bibliography of other work on this subject. If the phase velocity distribution of the region containing the N stations is known along with the instrument systems responses, Aki's method is readily adaptable to the present work. A technique for estimating the phase velocities based on cross spectrum analysis is now given.

As in section 3, the positions of the N stations are given in cartesian coordinates by (x_{μ}, y_{μ}) , $\mu = 2, 3, \dots, N$, the first station being at $(0, 0)$. The vector normal to the wave front of a strong signal makes an angle α with the x axis of the coordinate system. The signal, either natural in origin or artificially produced by high explosives, is assumed to be sufficiently strong to make the signal-to-noise ratio effectively infinite over the frequency range of interest. In this case, the signal observed at the μ th station will be

$$X_{\mu}(t) \cong S_{\mu}(t) = \int_{-\infty}^{\infty} h_{\mu}(u) S_1(t-u) du \quad (19)$$

As was seen in section 3, the cross spectral density of $S_{\mu}(t)$ and $S_{\nu}(t)$ is

$$f_{S_{\mu}S_{\nu}}(\omega) = e^{-i\Delta_{\mu\nu}(\omega)} f_{S_1}(\omega)$$

Thus, writing $f_{S_{\mu}S_{\nu}}(\omega) = C_{S_{\mu}S_{\nu}}(\omega) + iQ_{S_{\mu}S_{\nu}}(\omega)$, it is easily verified that

$$\Delta_{\mu\nu}(\omega) = -\arctan \frac{Q_{S_{\mu}S_{\nu}}(\omega)}{C_{S_{\mu}S_{\nu}}(\omega)} \quad (20)$$

Due to equations 19 and 20, a cross spectrum analysis of $X_{\mu}(t)$ and $X_{\nu}(t)$ will provide estimates $\hat{\Delta}_{\mu\nu}(\omega)$ of $\Delta_{\mu\nu}(\omega)$. For some cross spectrum programs [for example, Clark, 1958] the quantities $-\hat{\Delta}_{\mu\nu}(\omega)$ are part of the standard output.

Now, an easily computable estimate of the

phase velocity is obtained by averaging the values of $V(\omega)$ from (11) over all pairs of stations:

$$\hat{V}(\omega) = \frac{2\omega}{N(N-1)} \cdot \frac{\sum_{\mu=1}^N \sum_{\nu=\mu+1}^N (x_{\mu} - x_{\nu}) \cos \alpha + (y_{\mu} - y_{\nu}) \sin \alpha}{\hat{\Delta}_{\mu\nu}(\omega)}$$

The Rayleigh wave phase velocities are obtained by analyzing the records of the vertical instrument as usual. To obtain Love wave velocities for the frequency range covered after the arrival of Rayleigh waves, it may be necessary to correct the horizontal records for direction of propagation so that the transverse record may be analyzed.

5. APPLICATIONS

Description of data processing. The formulas of sections 3 and 4 will be applied to a system of three stations which were maintained during late 1959 and early 1960 by Sandia Corporation in the Manzano Mountains near Albuquerque, New Mexico. The stations were situated nearly in a straight line oriented in a northwesterly direction. The stations were numbered from northwest to southeast, and their separations in kilometers were $d_{12} = 3.9$, $d_{13} = 3.2$, and, thus, $d_{23} = 7.1$.

Each station was equipped with a Sprengnether vertical-motion seismometer which, when coupled with its amplifying system, had an effective passband of .05 to .15 cps (6.7- to 20-sec periods). Each station also contained a Benioff vertical-motion seismometer which, owing to the particular selections of filters in the recording system, had an effective passband restricted to the range of 1 to 4 cps (.25- to 1-sec periods).

The noise data in the long-period range was obtained in two 1-hour periods starting at 0056 and 2230 UT on December 11, 1959. The data were digitized at a sampling rate of one sample per second ($\Delta t = 1$ sec) which yielded 3001 observations ($n = 3000$) for each recording period. The quantities in parentheses are the values used in the design of the spectrum analysis in the notation of *Blackman and Tukey* [1958]. The Nyquist folding frequency was 0.5 cps, which, because of the 0.15-cps upper limit of the instrument passband, guarantees little or no aliasing of the spectrum.

Cross spectra for each pair of stations were obtained for both time periods by using 100 lags

($m = 100$), and the spectral estimates for two time periods were then averaged. This yielded spectral estimates at a frequency spacing of .005 cps ($\Delta f = .005$) with a resulting standard deviation based on an equivalent number of degrees of freedom equal to 58.

The short-period noise was obtained in a single 2.5-min time period beginning at 1700 UT on May 11, 1960. The sampling rate was 20 samples per second ($\Delta t = .05$ sec) which yielded 200 data points ($n = 3000$). The Nyquist folding frequency was 10 cps, which is again well above the instrument system passband.

Cross spectra for each pair of stations were obtained by using 50 lags ($m = 50$), which provides spectral estimates at a frequency spacing of 0.20 cps ($\Delta f = 0.20$) with an equivalent number of degrees of freedom equal to 59.

Since phase velocity data for Rayleigh waves were not available for the region containing the Sandia stations, a phase velocity distribution was taken from the literature for illustrative purposes. A distribution covering the frequency range of interest was given in Figure 9 of a paper by *Brune, Nafe, and Oliver* [1960] for a Russian nuclear explosion with epicenter near Novaya Zemlya recorded at Uppsala, Sweden. The values of $V(\omega)$ from this reference were used in the expression for $\theta_{\mu}(\omega)$ given in equation 4.

For the short-period range, $\theta_{\mu}(\omega)$ was assumed to be of the form given in equation 3 to conform with the nondispersive nature of body waves. For the epicentral distances of interest in the detection network, the travel times of body waves vary almost linearly with distance. It is seen from the *Seismological Tables* [Jeffrey, Bullen, 1948] that the slope of the travel time versus epicentral distance function is 1.7 sec/degree for P waves and 25.4 sec/degree for S waves. Let t_{ii} be the time in seconds for a given phase to travel from the i th to the j th station. Then, under the assumption that body waves travel parallel to the line of the stations, the following values are determined by the station spacings given earlier:

Wave Type	t_{12}	t_{13}	t_{23}
P	0.502	0.914	0.412
S	0.891	1.622	0.731

An interpretation of equations 7 and 15 in the simplicity, the case $N = 2$ will be considered.

discussion. The argument ω and subscripts are omitted where possible. The normalized cross spectral density of noise $c + iq$ defined in (10) can be represented in polar form by the expression $re^{i\phi}$, where $r = \sqrt{c^2 + q^2}$ and $\phi = \tan^{-1} q/c$. The quantity r is called the coefficient of coherence, having the property that $r = 0$ if the noises at the two stations are uncorrelated (at the given frequency) and $r = 1$ if they are perfectly correlated. The degree of noise correlation is then represented in increasing intensity by the value of r as it ranges from 0 to 1. The quantity ϕ represents the average phase lead of the noise at the second station relative to that at the first station.

In section 2, the change in the signal when the two stations may be described in complex form by the function $e^{i\theta}$, where θ now represents the phase lead of the signal at the second station relative to that at the first.

In the present notation, equations 7 and 15 can be written in the following form:

$$K = \left[\frac{1 + \cos \theta}{1 + r \cos \phi} \right]^{1/2} \quad (7')$$

$$K_M = \left[\frac{2}{1 + r \cos(\phi - \theta)} \right]^{1/2} \quad (15')$$

where $-\pi < \phi, \theta \leq \pi$.

Since θ and ϕ vary independently, K may assume any value in the range

$$0 \leq K \leq \sqrt{2/(1-r)}$$

The minimum value occurs when $\theta = \pi$, which corresponds to the case in which the signals at the two stations are completely out of phase. The maximum value occurs for $\theta = 0$ and $\phi = \pi$, which is the case when the signals are perfectly in phase and the noises completely out of phase. In this situation, an improvement in signal-to-noise ratio is obtainable for correlated noise which exceeds the value $\sqrt{2}$ achieved when the noises are uncorrelated. In fact, when $|\phi| < \pi$, the degree to which the noises are correlated when the two traces are added increases with r , and when $\theta \neq \pi$ and ϕ is near π the improvement ratio may become arbitrarily large. When the signals are perfectly in phase ($\theta = 0$) the improvement ratio is never smaller than $\sqrt{2/(1+r)}$, as is seen from equation 7'.

Thus, some improvement in signal-to-noise ratio will be obtained in this case provided only that the noises are not perfectly correlated. It is this fact that leads one to consider the signal correction technique described in section 4. The improvement factor for this method is given in the present notation by equation 15'. It is clear that K_M satisfies the inequality

$$\sqrt{2/(1+r)} \leq K_M \leq \sqrt{2/(1-r)}$$

The minimum value is assumed when the phase angles ϕ and θ are equal, and the maximum when $|\phi - \theta| = \pi$. For fixed r , K_M increases as $|\phi - \theta|$ increases from 0 to π . Thus, whether an improvement over the case of uncorrelated noise can be achieved depends critically on the differences in the phases of the signals and noises.

Let the first station of the pair be located at the origin of a cartesian coordinate system. The coordinates are so chosen that the second station is located at the point $(d, 0)$. If the normal vector to the signal wave front makes angle α with the x axis and if the noise is interpreted in an average sense to be a plane wave with normal vector making angle β with the x axis, then from (4) the phase difference is

$$\phi - \theta = \omega d \left(\frac{\cos \beta}{V_2} - \frac{\cos \alpha}{V_1} \right) \quad (21)$$

where V_1 and V_2 are the phase velocities of the signal and noise, respectively, and ω is the angular frequency. It is seen from this equation that the improvement factor K_M is enhanced by differences in signal and noise phase velocities and/or propagation angles. These effects will now be discussed in more detail and separately for the long-period and short-period data.

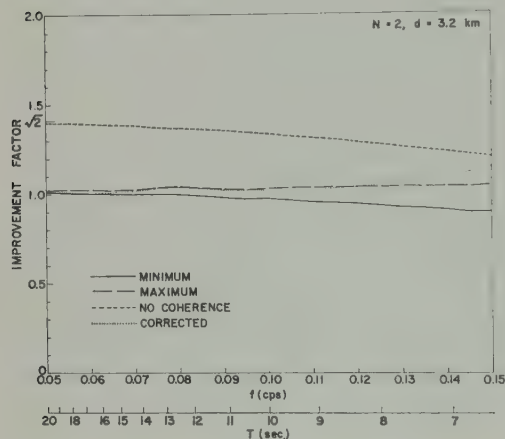
Discussion of results for the long-period data. The improvement factors for Rayleigh waves computed from equations 7 and 15 are plotted in Figures 1a-d. The curve labeled 'maximum' is a plot of the improvement factor for an uncorrected trace in the ideal case of no signal interference. This would correspond to a signal traveling at right angles to the line of stations ($\alpha = \pi/2$) in the presence of the observed background noise. The 'minimum' curve shows the improvement factor in the case in which the signal travels parallel to the line of stations ($\alpha = 0$). The curve labeled 'no coherence' demonstrates the effect of signal interference alone on the improvement factor when $\alpha = 0$.

It was obtained from equation 7 by setting $c_{N\mu N\nu}(\omega) = 0$ for all μ , ν , and ω . The curve marked 'corrected' demonstrates the improvement factor attained by using the modified technique of section 4 when the signal is assumed to travel parallel to the line of stations from northwest to southeast.

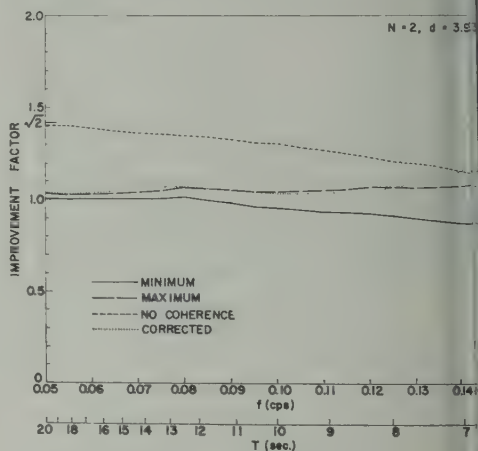
It is seen from Figures 1a-d that the signal-summing technique appears to provide little if any improvement in signal-to-noise ratio for Rayleigh wave signals. At least two conditions

contribute to this situation. First, noise recorded on the vertical instruments is basically a mixture of Rayleigh waves converging on the station from different directions [see, for example, *Gutenberg, 1958*]. Thus the phase velocities of signals and noises are the same, and, referred to equation 21, the difference in the direction of propagation will be the only feature which distinguishes the signal phase from that of the noise.

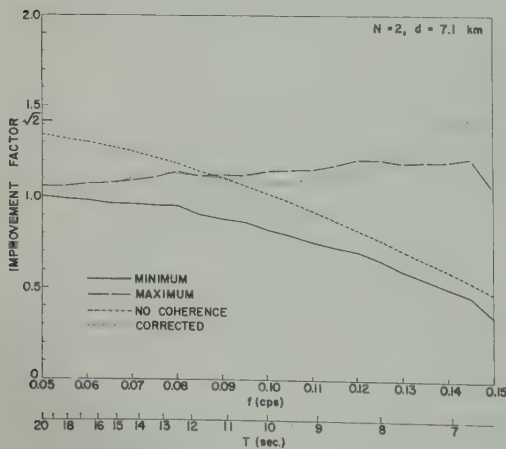
Second, on the day the long-period noises



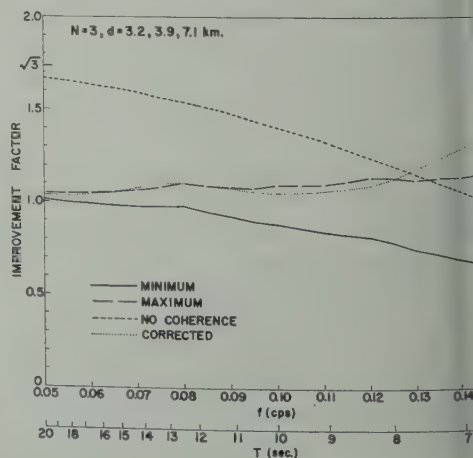
(a)



(b)



(c)



(d)

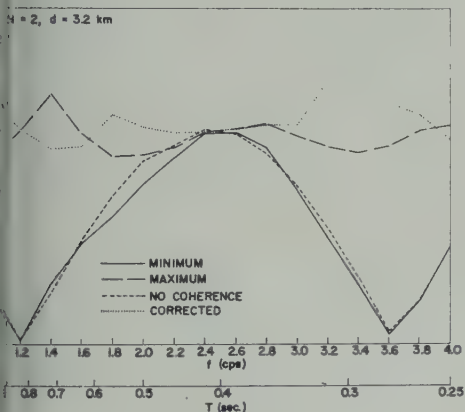
Fig. 1. a-d show the maximum and minimum improvement factors in the signal-to-noise ratio for an uncorrected signal, the minimum improvement factor in the signal-to-noise ratio for an uncorrected signal in uncorrelated noise (no coherence), and the improvement factor in the signal-to-noise ratio for a phase-corrected signal for various station separations d and numbers of stations N , where the signal is assumed to be a Rayleigh wave. Improvement factor = (signal-to-noise ratio of the summed signal) / (signal-to-noise ratio of the signal from a single station).

ed a strong directional noise polarization observed by *Hankins* [1960]. The direction of propagation was from northwest to south along the line of stations. This implies $\phi \cong 0$, so that, in the notation of the preceding subsection, the 'maximum' improvement factor will be approximately $[2/(1 + r)]^{1/2}$ and the 'minimum' $[(1 + \cos \theta)/(1 + r)]^{1/2}$, where $\theta = \omega d/V$ and $V_1 = V_2 = V$. r is nearly 1 for Rayleigh waves and is small over the given range of frequencies, so narrow bounds on the improvement factor are obtained on the graphs.

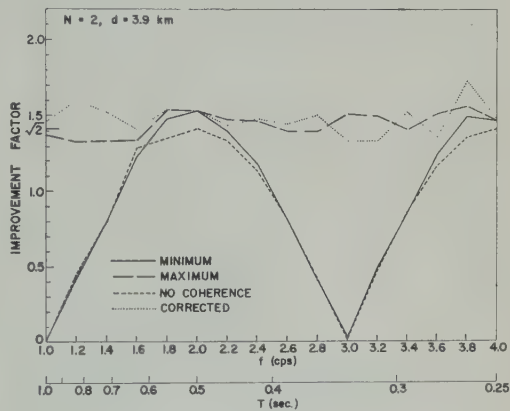
Since the signal was assumed to travel along

the line of the stations, $\phi = \theta$ and, as was discussed above, the minimum improvement for the modified signal-summing technique is attained. The sharp rise in the 'corrected' curves above 8-sec period is evidently due to the difference in crustal structure at Uppsala and Albuquerque which affects the values of V_1 and V_2 in this frequency range.

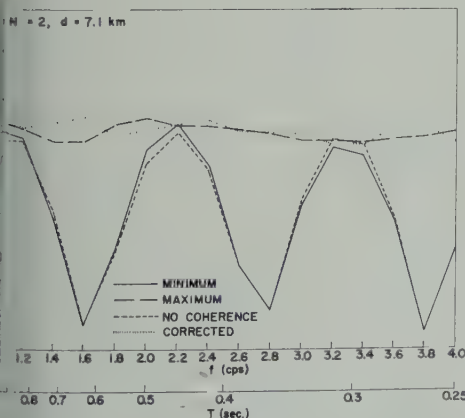
Discussion of results for the short-period data. The description of the curves in Figures 2a-d and 3a-d is the same as that for Figures 1a-d given in the preceding subsection. The characteristics of these curves will again be discussed for the case $N = 2$.



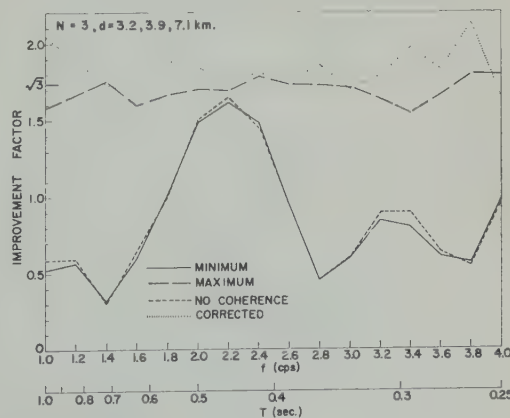
(a)



(b)



(c)



(d)

Fig. 2. a-d show the maximum and minimum improvement factors in the signal-to-noise ratio for an uncorrected signal, the minimum improvement factor in the signal-to-noise ratio for an uncorrected signal in uncorrelated noise (no coherence), and the improvement factor in the signal-to-noise ratio for a phase-corrected signal for various station separations d and numbers of stations N , where the signal is assumed to be a P wave.

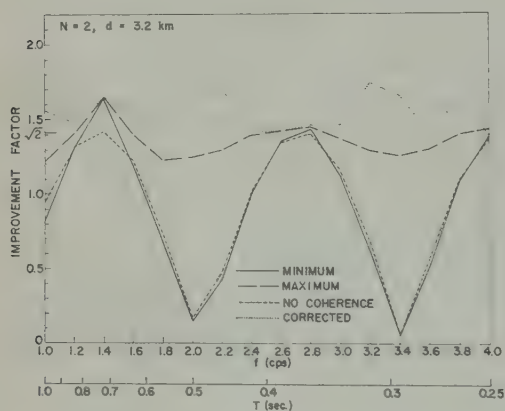
The noise correlation is small for this frequency range—the coefficient of coherence averages from 0.2 to 0.3 for all pairs of stations. For this reason the 'minimum' and 'no coherence' curves agree closely, as may be seen from equation 7'. The cyclic nature of these curves is due to the fact that $K \cong [1 + \cos \theta]^{1/2}$ and $\theta = \omega d/V$, passes through a number of multiples of 2π for the given range of frequencies ω .

Little variation is observed for the 'maximum' curves, since, taking $\theta = 0$ and r small in equation 7', K is restricted to a narrow range about $\sqrt{2}$. This is also the case for the 'corrected' curves; however it is noted that they frequently indicate a decided increase in the improvement

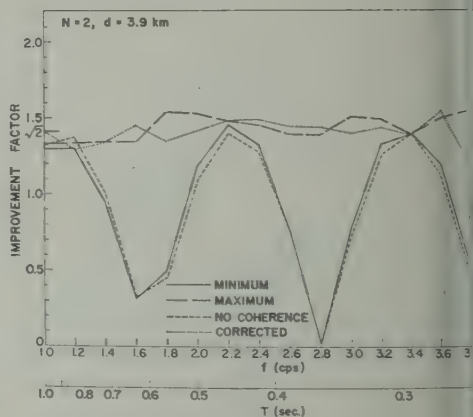
factor for the modified technique over the achievable for the unmodified procedure. This may be attributed primarily to the difference in phase velocities for signals and noises. If signals were assumed to be body waves, hence to have apparent velocities greater than the phase velocities of the noises which are of surface waves.

Conclusions and extensions. It may tentatively be concluded that the signal-summing technique is a useful method for improving the rms signal-to-noise ratio for seismic events.

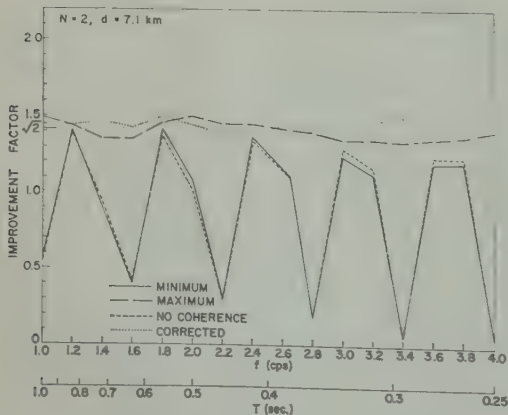
In the short-period range the signal consists of body waves, and the technique provides a maximum improvement comparable to that obtained



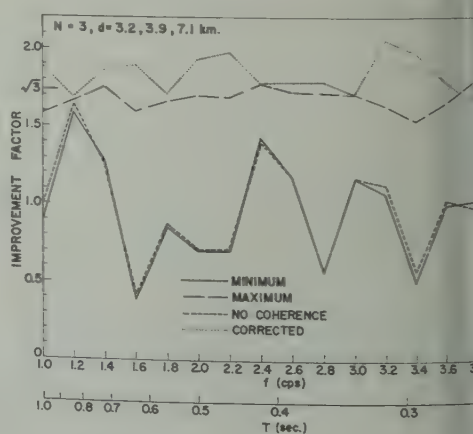
(a)



(b)



(c)



(d)

Fig. 3. *a-d* show the maximum and minimum improvement factors in the signal-to-noise ratio for an uncorrected signal, the minimum improvement factor in the signal-to-noise ratio for an uncorrelated noise (no coherence), and the improvement factor in the signal-to-noise ratio for a phase-corrected signal for various station separations d and numbers of stations N , where the signal is assumed to be an S wave.

or noninterfering signals in uncorrelated. However, to guarantee this improvement, modified signal-summing technique must be used. For the station spacings given in the table, the expected increase in the improvement factor with an increase in number of stations is realized.

At the lower frequencies where the signal is a surface wave, the correction procedure is satisfactory in order to take advantage of the directional properties of the signal. It is believed that the poor showing of the method exhibited in Figures 1a-d is due to an unfortunate content of the noise with the assumed signal. Comparing Figures 1a-c with Figure 1d, we see no increase in the improvement factor with the number of stations. This is also probably due to the directional polarization of the noise, which may be due to inadequate station spacings. Further investigations under different conditions and greater station separations are needed for a full evaluation of the technique in the frequency range.

Equations 7 and 15 may be generalized to take into account the possibility of signal attenuation and different noise spectra for different stations at the expense of somewhat more complicated mathematical formulas. It would be necessary to know the form of the attenuation factors in this case. The generalized formulas could be applied to the study of the signal-summing technique for widely spaced stations as the manned stations of a detection network. Here, the traces from all stations containing a 'suspect' signal would be transmitted to a master location for analysis. Again it is noted that, under favorable conditions of noise reduction, an improvement in the signal-to-noise ratio for the long periods over that indicated in Figures 1a-d would be obtained. An extensive study of this application would be warranted by the fact that, once a network is established, the signal-summing technique can be applied at only a small additional cost.

Acknowledgments. I am indebted to Dr. C. F. B. Vice President of Research, Sandia Corporation, for the suggestion that initiated this research. Thanks are also due Drs. B. F. Murphy, J. Norris, and W. D. Weart, as well as Dr. J. for several helpful comments. This work was performed under the auspices of the U. S. Atomic Energy Commission.

REFERENCES

- Aki, K., Study of earthquake mechanism by a method of phase equalization applied to Rayleigh and Love waves, *J. Geophys. Research*, **65**, 729-740, 1960.
- Blackman, R. B., and J. W. Tukey, The measurement of power spectra from the point of view of communications engineering, *Bell System Tech. J.*, **37**, 185-282 and 485-568, 1958. (Republished by Dover Publications, New York, 1959.)
- Brune, J. N., J. E. Nafe, and J. E. Oliver, A simplified method for the analysis and synthesis of dispersed wave trains, *J. Geophys. Research*, **65**, 287-304, 1960.
- Clark, N., Tukey spectrum estimation program (CS Tuks), *Share distribution 574*, Convair, San Diego, 1958.
- Doob, J. L., *Stochastic Processes*, John Wiley & Sons, New York, 1953.
- Frank, H. R., and W. E. N. Doty, Signal-to-noise ratio improvements by filtering and mixing, *Geophysics*, **18**, 587-604, 1953.
- Goodman, N. R., On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two-dimensional stationary gaussian process, *Sci. Paper 10*, Engineering Statistics Laboratory, New York University, 1957.
- Grenander, U., and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley & Sons, New York, 1957.
- Gutenberg, B., Microseisms, *Advances in Geophysics*, **5**, 53-92, 1958.
- Hankins, D. M., Noise study using visicorder playbacks of signals received on long-period seismometer at Sandia, *Sandia Corp. Tech. Mem. SCTM 36-60(51)*, 1960.
- Holzmann, F. M., Frequency theory of the grouping of signals on a background of correlated noises, *Izvest. Akad. Nauk SSSR, Ser. Geofiz.*, 769-780, 1960.
- Jeffreys, H., and K. E. Bullen, *Seismological Tables*, London Office of the British Association, Burlington House, W. 1, 1948.
- Kogan, S. O., I. P. Pasechnik, and D. D. Sultanov, The difference in the periods of seismic waves generated by subsurface explosions and those generated by earthquakes, *Doklady Akad. Nauk SSSR*, **129**, 1283-1286, 1959.
- Tukey, J. W., The sampling theory of power spectrum estimates, *Symposium on Applications of Autocorrelation Analysis to Physical Problems*, Woods Hole, Massachusetts, Office of Naval Research, Washington D. C., pp. 47-67, 1949.
- Wiener, N., Generalized harmonic analysis, *Acta Math.*, **55**, 117-258, 1930.
- Young, D. A., A Fortran II subroutine system to perform various operations on a time series using the IBM-704 digital computer, *Sandia Corp. Tech. Mem. SCTM 71-61(54)*, 1961.

(Manuscript received July 15, 1961; revised August 29, 1961.)

Free Oscillations of the Moon

H. TAKEUCHI AND M. SAITO

Geophysical Institute, Tokyo University, Tokyo

N. KOBAYASHI

Department of Precision Mechanics, Chūo University, Tokyo

Abstract. Theoretical studies are made on the free oscillation of a homogeneous self-gravitating elastic sphere. Numerical results for torsional and spheroidal oscillations are shown in Tables 1 and 3, respectively. In these tables, are shown also the free periods of a model moon, for which density $\rho = 3.33$ g/cm³ and shear wave velocity $V_s = 4.7$ km/sec. For the model moon, the effect of gravity on the periods of oscillations is shown to be negligible. From Table 3, we can make an estimation of the effect of curvature on the phase velocity dispersion.

Oscillations of the earth have recently been studied theoretically and verified experimentally [e.g., Pekeris, Alterman, and Jarosch, 1960]. The moon will be the next experimental object of this kind of study. It is thus desirable to make theoretical studies of the free oscillations of the moon. As is well known [Jeffreys, 1939, p. 151], the moon is a sphere of radius 1738 km and mean density 3.33 g/cm³. Hydrostatic pressure at its center is about 50000 atm, greater than that at the 20° discontinuity within the earth. Neither polymorphic transition nor appreciable increase in density by self-compression will be taking place within the moon. According to Jeffreys, if entirely of one material, the density of the moon would range from 3.28 at the surface to 3.41 at the center. Thus, in the first approximation, we may assume the moon to be a homogeneous elastic sphere. The rigidity of the moon is the most unknown quantity in studying the problem of its free oscillations. The density, 3.33 g/cm³, will suggest the basic material immediately under the Proterozoic discontinuity as a moon material. In the following calculation, we shall tentatively assume the shear wave velocity within the moon to be 4.7 km/sec and assume $\lambda = \mu$. The corresponding rigidity of the moon will be 0.7356 $\times 10^{12}$ dyne/cm².

Possible frequencies of torsional oscillations of a homogeneous elastic sphere are determined by the equation [Love, 1944, p. 284]

$$\frac{n-1}{ka} + \frac{\psi_n'(ka)}{\psi_n(ka)} = 0 \quad (1)$$

where

$$k^2 a^2 = \frac{\rho P^2 a^2}{\mu}, \quad (2)$$

$$\psi_n(x) = (-1)^n \sqrt{\frac{\pi}{2}} \frac{J_{n+0.5}(x)}{x^{n+0.5}}$$

In Table 1 are given the values of $P = pa\sqrt{\rho/\mu} = 2\pi a/TV_s$ and $C/V_s = P/n + 0.5$ obtained by solving (1) for $n = 2$ to 10 together with the corresponding free periods T when $a = 1738$ km, $\rho = 3.33$ g/cm³, and shear wave velocity $V_s = \sqrt{\mu/\rho} = 4.7$ km/sec. C in Table 1 is the phase velocity of the surface wave corresponding to the torsional oscillation. In the last column of Table 1 are shown approximate values of P obtained by the variational calculus method developed in a previous paper [Takeuchi, 1959]. In the last step of the variational calculus, we have only to solve a symmetric determinant equation for P^2 , the elements of which are shown

TABLE 1. Computations for Torsional Oscillations

n	P	C/V_s	T , min	P (approx.)
2	2.501	1.00	15.48	2.519
3	3.865	1.104	10.02	3.886
4	5.094	1.132	7.60	5.107
5	6.266	1.139	6.18	6.269
6	7.404	1.139	5.23	7.403
7	8.520	1.136	4.55	8.523
8	9.621	1.132	4.02	9.634
9	10.71	1.127	3.62	10.736
10	11.79	1.123	3.28	11.833

TABLE 2. Scheme for Showing Elements of the Determinant

Element	P^2	N	
2 ² 2	0.2	0.33333333	-0.66666666
2 ⁶ 6	0.031111111	0.042857143	-0.085714286
6 ⁶ 6	0.0253040293	0.0268055268	0.0250349650
2 ² 20	0.0240045767	0.0245568467	-0.0291136933
6 ⁶ 20	0.0276532327	0.0286369771	0.0264777328
20 ² 20	0.0214971602	0.0216301740	0.0298052973

in Table 2. The full form of the element 2²2, for example, is as follows:

$$2^22 = 0.2P^2 - (0.3N - 0.6) \tag{3}$$

where $N = n(n + 1)$. Table 1 shows that the variational calculus method gives very good approximations for the nondimensional frequency P . For our moon model, the longest period in the torsional oscillations is about 15.5 min. When the rigidity of the moon is different from that assumed above, we can get the corresponding period by using the result that the period is proportional to the square root of the rigidity.

Possible frequencies of spheroidal oscillations of a homogeneous elastic sphere are determined by the equations [Love, 1926, p. 136]

$$\frac{ng\rho/a}{\mu\alpha^2 - p^2\rho} \left\{ \frac{\alpha^2 a^2}{n} - 2 \frac{n^2 - 1}{n} + \nu + \frac{2}{n} \alpha a \frac{\psi_n'(\alpha a)}{\psi_n(\alpha a)} \right\} + 2(n - 1) + \nu + 2\alpha a \frac{\psi_n'(\alpha a)}{\psi_n(\alpha a)}$$

$$\alpha^2 a^2 \left[\frac{ng\rho/a}{\mu\alpha^2 - p^2\rho} \left\{ 1 + 2 \frac{n + 2}{\alpha a} \frac{\psi_n'(\alpha a)}{\psi_n(\alpha a)} \right\} + \frac{\lambda + 2\mu}{\mu} + 2 \frac{n + 2}{\alpha a} \frac{\psi_n'(\alpha a)}{\psi_n(\alpha a)} \right]$$

$$= \frac{ng\rho/a}{\mu\delta^2 - p^2\rho} \left\{ \frac{\delta^2 a^2}{n} - 2 \frac{n^2 - 1}{n} + \nu + \frac{2}{n} \delta a \frac{\psi_n'(\delta a)}{\psi_n(\delta a)} \right\} + 2(n - 1) + \nu + 2 \delta a \frac{\psi_n'(\delta a)}{\psi_n(\delta a)}$$

$$\delta^2 a^2 \left[\frac{ng\rho/a}{\mu\delta^2 - p^2\rho} \left\{ 1 + 2 \frac{n + 2}{\delta a} \frac{\psi_n'(\delta a)}{\psi_n(\delta a)} \right\} + \frac{\lambda + 2\mu}{\mu} + 2 \frac{n + 2}{\delta a} \frac{\psi_n'(\delta a)}{\psi_n(\delta a)} \right]$$

$$\nu = \frac{6(n^2 - 1)}{\{(2n + 1)q - 2n(n - 1)\}}, \quad q = \frac{\rho p^2 a^2}{\mu}$$

$$\frac{gp/a}{\mu\alpha^2 - p^2\rho} = \frac{\rho ga}{\alpha^2 a^2 - \rho p^2 a^2} \tag{4}$$

Two roots x^2 of the following equation

$$\frac{\lambda + 2\mu}{\mu} \left(\frac{\rho p^2 a^2}{\mu} \right)^2 x^4 - \left\{ \left(\frac{\lambda + 3\mu}{\mu} \right) \frac{\rho p^2 a^2}{\mu} + 4 \frac{\rho ga}{\mu} \right\} \frac{\rho p^2 a^2}{\mu} x^2 + \left\{ \left(\frac{\rho p^2 a^2}{\mu} \right)^2 + 4 \frac{\rho ga}{\mu} \frac{\rho p^2 a^2}{\mu} - n(n + 1) \left(\frac{\rho ga}{\mu} \right)^2 \right\} = 0$$

give the coefficient in the equation $\alpha^2 a^2$ or

$$\delta^2 a^2 = \frac{\rho p^2 a^2 x^2}{\mu}$$

g in (4) and (5) is the gravity at the surface of the sphere. For a nongravitating sphere, we have $x^2 = 1$ and $1/3$ from (5) and $\nu = 0$ in (4). The left- and right-hand sides of the first equation in (4) are referring to $\alpha^2 a^2$ ($x^2 = 1$) and $\delta^2 a^2$ ($x^2 = 1/3$), respectively. On the left-hand side of the equation, the terms independent of g , $gp/a/\mu\alpha^2 - p^2\rho$ become much smaller than the terms dependent on it, whereas on the right-hand side we have the contrary. In Table 3 are shown values of $P = \rho a \sqrt{\rho/\mu} = 2\pi a/TV$, $C/V_s = P/n + 0.5$ for a nongravitating sphere. The problem thus formulated together with the corresponding free periods T for a moon model are in Table 1. In (4) and (5), the effect of gravity on the free spheroidal oscillations is represented by a nondimensional parameter $\rho ga/\mu$. For

moon model used in Tables 1 and 3, we have $\rho ga/\mu = 0.1272$, whereas for an earth model $\rho ga/\mu = 4.217$ [Love 1926, pp. 142, 143] in which $\rho = 5.52 \text{ g/cm}^3$, $\mu = 0.819 \times 10^{12} \text{ dyne/cm}^2$, $a = 6370 \text{ cm}$, $g = 982.4 \text{ cm/sec}^2$, we have $\rho ga/\mu = 4.217$. Equations 4 and 5 give $P = 2.872, 4.000, 5.022$ for $n = 2, 3$, and 4 , respectively, $\rho ga/\mu = 4.217$ and $P = 2.655$ for $n = 2$ and $\rho ga/\mu = 0.1272$. Comparing these results with those in Table 3, we see that for the above

TABLE 3. Computations for Spheroidal Oscillations

P	C/V_*	T , min.
2.639	1.056	14.67
3.916	1.119	9.89
5.009	1.113	7.73
6.033	1.097	6.42
7.023	1.080	5.51
7.995	1.066	4.84
8.956	1.054	4.32
9.908	1.043	3.91
10.855	1.034	3.57
16.477	0.9986	2.35

the relative decrease of periods of spheroidal oscillations is less than 1 per cent for n than 4, and for the moon model the decrease is negligible even for $n = 2$. Thus we consider the periods in Table 3 as those for a rotating moon. The decrease of periods for $n = 2$ for the earth model has already been noted by Love [1926, p. 143] and is frequently referred to as being a decrease of from 66 min to 64 min in the problem of the earth's free oscillations.

There is another point to be noted in Table 3 (also in Table 1). The phase velocity C in Table 3 increases as n decreases. Because the moon is assumed to be homogeneous, the wave motion is entirely due to the spherical curvature. For $n \rightarrow \infty$, C will be equal to the Rayleigh velocity $0.9194 V_*$. Thus 10 per cent and

20 per cent increase in the phase velocity are attained at $n = 15$ and 5, respectively. The corresponding wavelengths are 0.41 and 1.14 times the radius of the sphere, respectively. Similar values are obtained in the corresponding problem of a circular cylinder [Ewing, Jardetzky, and Press, 1957, p. 265].

Note added in proof. After the completion of our work, a paper by Bolt [1961] on spheroidal oscillations of the moon was brought to our attention. Our model moon is nothing but his model 1. In comparison with his paper, our calculations include larger n , torsional oscillations, and curvature and gravity effects.

REFERENCES

- Bolt, B. A., Spheroidal oscillations of the moon, *Nature*, 188, 1176-1177, 1961.
 Ewing, M., W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media*, McGraw-Hill Book Co., New York, 1957.
 Jeffreys, H., *The Earth*, Cambridge University Press, 1952.
 Love, A. E. H., *Some Problems of Geodynamics*, Cambridge University Press, 1926.
 Love, A. E. H., *Theory of Elasticity*, Dover, New York, 1944.
 Pekeris, C. L., Z. Alterman, and H. Jarosch, Comparison of theoretical with observed values of the periods of free oscillation of the earth, *Proc. Natl. Acad. Sci. U. S.*, 47, 91-98, 1961.
 Takeuchi, H., Torsional oscillations of the earth and some related problems, *Geophys. J.*, 2, 89-100, 1959.

(Manuscript received June 14, 1961.)

Refraction Arrivals along a Thin Elastic Plate Surrounded by a Fluid Medium¹

J. H. ROSENBAUM

*Shell Development Company, Exploration and Production Research Division
Houston, Texas*

Abstract. An asymptotic solution is derived for the first arrival of significant amplitude and low frequency, which is refracted along a thin high-velocity plate surrounded by a fluid medium. This first arrival travels with approximately the velocity of a longitudinal plate wave in the refracting layer. The shape of the signal depends upon the contrast (in density and elastic constants) between the plate and the fluid as well as upon the distance (in units of plate thickness) that the signal has been refracted along the plate. Solutions in closed form are given. For the case of moderate contrast or large distance, we treat the problem of a point and of a line source in the fluid emitting a step (or δ) pressure pulse. The refracted signal is approximately a gaussian curve (or its derivative). For the case of large contrast and moderate distance, we treat the problem of a line source in the fluid emitting a step pressure pulse. The refracted signal is described by an Airy function. For intermediate cases, the refracted signal for the line source for the point source is given in the form of an integral, which must be evaluated numerically. Because of the relatively simple nature of the results, interesting features of the signal are readily derived.

In an appendix, we discuss briefly a thin high-velocity plate shallowly submerged in a liquid half-space. The low-frequency refracted signal from a step pressure line source is described by an Airy function for all cases of contrast.

Introduction. In the theory of wave propagation in layered elastic media, the form and amplitude of a signal refracted through a thin layer whose compressional (P) wave velocity is greater than that of the surrounding medium, is of great interest. The elastic plate in a fluid medium is a model sufficiently simple to permit analytical analysis, but it shows many important features of a first significant arrival through a 'high-velocity' layer. By significant arrival we mean a signal of sufficient amplitude to be observable and of period low compared with the vertical transit time of P waves through the layer.

Figure 1 illustrates the model under consideration. A point source of explosive sound is located in a fluid of sound velocity c_1 and density ρ_1 at a distance d above the upper surface of an elastic plate of thickness H , compressional (P) wave velocity c_3 , rotational (S) wave velocity c_2 , and density ρ_2 . A detector is located at a vertical distance z above the upper surface of the plate and at a horizontal distance r from the source. The pressure signal P_0 , emitted by the source, is described in the sur-

rounding fluid by the equation

$$P_0 = \frac{A}{R} H_u \left[t - \frac{R}{c_1} \right] \exp \left[-\frac{t - R/c_1}{\theta} \right] \quad (1)$$

where t is the time variable, H_u the Heaviside unit step function, θ the exponential decay time constant, A an amplitude factor, and $R^2 = r^2 + (z - d)^2$.

The pressure response, P , at the detector is then given by the equation

$$P = P_0 + \frac{\bar{A}}{\pi} \operatorname{Re} \int_0^\infty \frac{e^{i\bar{a}\tau} d\bar{\omega}}{(1/\bar{\theta}) + i\bar{\omega}} \cdot \int_0^\infty \frac{e^{-a_1(\bar{a} + \bar{s})}}{a_1} J_0(K\bar{r}) K \frac{\Delta + W}{\Delta} dK \quad (2)$$

where

$$\Delta = \Delta_1 \cdot \Delta_2$$

$$\Delta_1 = \frac{\rho_1}{\rho_2} \bar{\omega}^4 \frac{a_2}{a_1} \sinh \frac{a_2}{2} \sinh \frac{a_{22}}{2}$$

$$- 4K^2 a_2 a_{22} \sinh \frac{a_2}{2} \cosh \frac{a_{22}}{2}$$

$$+ (2K^2 - \bar{\omega}^2)^2 \cosh \frac{a_2}{2} \sinh \frac{a_{22}}{2}$$

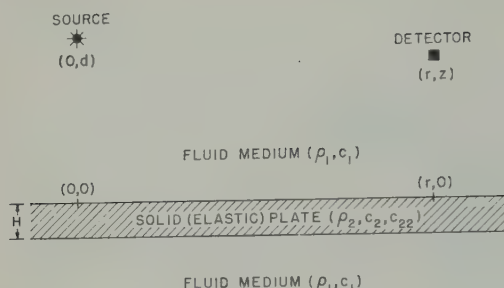


Fig. 1. Refraction through a thin, high-velocity plate in a fluid medium.

$$\Delta_a = \frac{\rho_1}{\rho_2} \bar{\omega}^4 \frac{a_2}{a_1} \cosh \frac{a_2}{2} \cosh \frac{a_{22}}{2} - 4K^2 a_2 a_{22} \cosh \frac{a_2}{2} \sinh \frac{a_{22}}{2} + (2K^2 - \bar{\omega}^2)^2 \sinh \frac{a_2}{2} \cosh \frac{a_{22}}{2}$$

$$W = \frac{\rho_1}{2\rho_2} \bar{\omega}^4 \frac{a_2}{a_1} \left[4K^2 a_2 a_{22} \sinh a_2 \cosh a_{22} - (2K^2 - \bar{\omega}^2)^2 \cosh a_2 \sinh a_{22} - \frac{\rho_1}{\rho_2} \bar{\omega}^4 \frac{a_2}{a_1} \sinh a_2 \sinh a_{22} \right]$$

$$a_1 = \left[K^2 - \frac{c_{22}^2}{c_1^2} \bar{\omega}^2 \right]^{1/2}$$

$$a_2 = \left[K^2 - \frac{c_{22}^2}{c_2^2} \bar{\omega}^2 \right]^{1/2}$$

$$a_{22} = (K^2 - \bar{\omega}^2)^{1/2}$$

All variables have been normalized with respect to the thickness of the plate, H , and the vertical transit time of rotational (S) waves through the plate, H/c_{23} . Specifically, $\bar{r} = r/H$, $\bar{d} = d/H$, $\bar{z} = z/H$, $\tau = tc_{23}/H$, $\bar{\omega} = \omega H/c_{23}$, $K = kH$, $\bar{A} = A/H$, and $\bar{\theta} = \theta c_{23}/H$. ω is the angular frequency and k the horizontal wave number. When K is real and positive, the square root a_1 is real and positive for imaginary values of $\bar{\omega}$. $\Delta_s = 0$ is the period equation for the so-called symmetric modes; $\Delta_a = 0$ is the period equation for the antisymmetric modes.

If the horizontal distance r is sufficiently large compared with the plate thickness H , and if r is also sufficiently large compared with the vertical distance $d + z$, then we have essentially a 'thin' plate and can observe early refraction

arrivals. The response at the detector will be best described as an interference phenomenon and the use of a modal type of analysis is indicated. This method of analysis is especially useful here since, as we shall see, the observed signal is due mainly to contributions from a single mode. The method which we have used is similar to that described by Rosenbaum [1]. Here we shall only mention that for a sufficiently thin plate the first 'significant' arrival is closely connected with the longitudinal plate wave. As a matter of fact, an investigation of the roots of the period equation $\Delta = 0$ reveals that $\Delta_s = 0$ has a root for small real $\bar{\omega}$ and K which is associated with horizontal group and phase velocities greater than the velocity of sound in the fluid. This root is a solution to the equation $\Delta_s = 0$ in the limit $\bar{\omega} \rightarrow 0$ as $K \rightarrow 0$ so that $d\bar{\omega}/dK \rightarrow \bar{\omega}/K \rightarrow 2[1 - (c_{23}^2/c_2^2)]^{1/2}$. In this limit, the solution represents a decoupled oscillation, since the motion of the plate surface is entirely horizontal. This is shown also by the fact that the function W , and consequently the numerator of the integrand in (2), vanishes in this limit. The root discussed here constitutes the end point of a complex branch of roots of the equation $\Delta_s = 0$, as is indicated by the fact that $d^2\bar{\omega}/dK^2$ is imaginary positive at $K = 0$. By expanding our solution around this end point, we can obtain expressions in simple analytic form for the first 'significant' refracted arrival. Our solution is the first term of an asymptotic series in inverse powers of the length of the refraction path. There are two separate reasons why the end-point contribution yields an expression for the first refracted arrival. If the contrast in density and elastic constants between the plate and the surrounding medium is moderate, then $|d^2\bar{\omega}/dK^2|$ is sufficiently large so that additional contributions, if any, from the preceding mode are quite small. If the contrast in density and elastic constants is large, then $|d^2\bar{\omega}/dK^2|$ is small, but $d^3\bar{\omega}/dK^3$ is real and negative. Additional contributions, if any, to the pressure signal for this mode must therefore arrive at the detector subsequent to the 'end-point' contribution. It is important to note that within a certain distance around the origin of the K plane where $\bar{\omega}$ and K have convergent Maclaurin expansions, several points of the integral over K for the pre-

² The longitudinal plate wave is the two-dimensional analog of the longitudinal rod wave.

are on paths of steepest descent from the point at the origin. An evaluation of the integral in (2) by the method of stationary phase in the saddle-point method in the usual way is therefore not possible.

The algebraic manipulation leads to the following results for the solution of the period when $\Delta_s = 0$ in the vicinity of the end point $\bar{\omega} = 0$ and $K = 0$, expressed in the form of $\bar{\omega}$ as a function of K :

$$\bar{\omega}_0' = 2[1 - (c_{22}^2/c_2^2)]^{1/2} \quad (3)$$

$$\bar{\omega}_0'' = -\frac{\rho_1 \bar{\omega}_0' [1 - (2c_{22}^2/c_2^2)]^2}{2\rho_2 a_{10}'} \quad (4)$$

$$a_{10}' = i \left[\frac{c_{22}^2}{c_1^2} (\bar{\omega}_0')^2 - 1 \right]^{1/2}$$

$$a_{10}'' = -\frac{(c_{22}^2/c_1^2) \bar{\omega}_0' \bar{\omega}_0''}{a_{10}'}$$

$$= -\frac{1}{4} \bar{\omega}_0' \left[1 - \frac{2c_{22}^2}{c_2^2} \right]^2$$

$$= \frac{3(\bar{\omega}_0'')^2 [7(c_{22}^2/c_2^2)(\bar{\omega}_0')^2 - 3]}{4\bar{\omega}_0' [1 - (2c_{22}^2/c_2^2)]^2} - \frac{3\bar{\omega}_0'' a_{10}''}{2a_{10}'} \quad (5)$$

$$a_{10}''' = -\frac{(c_{22}^2/c_1^2) \bar{\omega}_0' \bar{\omega}_0'''}{a_{10}'} \quad (6)$$

where the primes designate derivatives with respect to K and the subscript 0 refers to the evaluation of a function at the end point $K = 0$. Note that the group velocity $\bar{\omega}_0'$ and the phase velocity $\bar{\omega}/K$ are both equal to the longitudinal wave velocity in the limit when $K = 0$. Note that when $\rho_2 \gg \rho_1$ and $\bar{\omega}_0' \gg c_1/c_{22}$, and $|a_{10}''|$ become very small, and $\bar{\omega}_0'''$ is finite.

We need not go into the details of the analysis, since the essential features have been described elsewhere [Rosenbaum, 1960], but shall outline the general results. Since the case of moderate contrast, or of sufficiently long reentrant path along the plate if the contrast is small, is most interesting, we shall discuss it in some detail.

General solution. We first consider the case in which the point source emits a Heaviside pressure pulse. This is equivalent to letting $\bar{\omega} \rightarrow 0$ in (2) and is a fair approximation for

the case $\theta \gg H/(c_{22}\bar{\omega}_0')$. We change the order of integration and integrate over $\bar{\omega}$. The contribution of interest to us is given by the residue at a simple pole, which is due to the particular root of $\Delta_s = 0$ discussed above. This root and the contribution of the pole to the integral are evidently functions of K ; the behavior of the root in the vicinity of $K = 0$ is described approximately by (3) and (5). We use the values 3 to 6 in our expansion of the integrand around $K = 0$ and obtain the following approximate early pressure response at the detector:

$$\begin{aligned} & \frac{\bar{A} |\bar{\omega}_0''|}{2 |a_{10}'| \bar{\omega}_0' (\pi \bar{b})^{1/2}} \\ & \cdot \int_0^\infty \left\{ \cos \left[(\bar{x} - \bar{\omega}_0' \tau) \bar{b}^{-1/3} \xi + \frac{\xi^3}{3} \right] \right. \\ & \left. + \sin \left[(\bar{x} - \bar{\omega}_0' \tau) \bar{b}^{-1/3} \xi + \frac{\xi^3}{3} \right] \right\} \\ & \times \xi^{1/2} \exp(-\bar{a} \bar{b}^{-2/3} \xi^2) d\xi \quad (7) \end{aligned}$$

where

$$\bar{x} = \bar{r} + |a_{10}'| (\bar{d} + \bar{z})$$

$$\bar{a} = [|\bar{\omega}_0''| \tau - |a_{10}''| (\bar{d} + \bar{z})]/2$$

and

$$\bar{b} = -[\bar{\omega}_0''' \tau + i a_{10}''' (\bar{d} + \bar{z})]/2$$

In (7), as well as in all subsequent expressions, we have neglected contributions of exponentially smaller order, which may be ascribed to waves traveling horizontally toward the source.

Similarly, we can obtain the early pressure response for the case in which the source emits a δ -function pressure pulse. This is equivalent to letting $\bar{\theta} \rightarrow 0$ and $\bar{A} \rightarrow \infty$ in (2) in such a way that $\bar{A} \bar{\theta} \rightarrow \bar{G}$, a constant which describes the strength of the source. The solution may be a fair approximation for the case $\theta \ll H/(c_{22}\bar{\omega}_0')$. For the early pressure response at the detector, we obtain

$$\begin{aligned} & \frac{\bar{G} |\bar{\omega}_0''|}{2 |a_{10}'| (\pi \bar{r})^{1/2} \bar{b}^{5/6}} \\ & \cdot \int_0^\infty \left\{ \sin \left[(\bar{x} - \bar{\omega}_0' \tau) \bar{b}^{-1/3} \xi + \frac{\xi^3}{3} \right] \right. \\ & \left. - \cos \left[(\bar{x} - \bar{\omega}_0' \tau) \bar{b}^{-1/3} \xi + \frac{\xi^3}{3} \right] \right\} \\ & \times \xi^{3/2} \exp(-\bar{a} \bar{b}^{-2/3} \xi^2) d\xi \quad (8) \end{aligned}$$

It is often convenient to solve our problem for the two-dimensional case of a line source and a line receiver. The appropriate expression for such a case is obtained by the substitution of $2H \cos(K\bar{r})$ for $J_0(K\bar{r})K$ in (2). We note here the fact that for large \bar{r} , $[(2K)/(\pi\bar{r})]^{1/2} \cos[K\bar{r} - (\pi/4)]$ is a sufficiently good approximation to $J_0(K\bar{r})K$ in (2). This is precisely the reason we can write the results (7) and (8).³ When the line source emits a Heaviside step pressure pulse, we obtain for the early response at the detector

$$\frac{A |\bar{\omega}_0''|}{|a_{10}'| \bar{\omega}_0' \bar{b}^{1/3}} \int_0^\infty \cos \left[(\bar{x} - \bar{\omega}_0' \tau) \bar{b}^{-1/3} \xi + \frac{\xi^3}{3} \right] \cdot \exp(-\bar{a} \bar{b}^{-2/3} \xi^2) d\xi \quad (9)$$

whereas for the δ -function source, we obtain

$$\frac{G' |\bar{\omega}_0''|}{|a_{10}'| \bar{b}^{2/3}} \int_0^\infty \sin \left[(\bar{x} - \bar{\omega}_0' \tau) \bar{b}^{-1/3} \xi + \frac{\xi^3}{3} \right] \cdot \exp(-\bar{a} \bar{b}^{-2/3} \xi^2) \xi d\xi \quad (10)$$

Since the change from three to two dimensions is accomplished by an integration over one space coordinate, \bar{A} and \bar{G} of (7) and (8) have been replaced by A and $G' = A \bar{\theta}$.

Moderate contrast or long refraction path. For this case $|\bar{a}|^{3/2}$ is sufficiently large compared with $|\bar{b}|$, and hence we can ignore the third-order terms in our expansion around $K = 0$. The integrals over K can now be expressed in terms of tabulated functions. We obtain for the point source emitting a Heaviside step pressure pulse the approximate early pressure response at the detector

$$\frac{\bar{A} |\bar{\omega}_0''| (\pi)^{1/2}}{8 |a_{10}'| \bar{\omega}_0' \bar{r}^{1/2} \bar{a}^{3/4}} \eta^{3/2} \left[I_{-3/4} \left(\frac{\eta^2}{2} \right) + I_{-1/4} \left(\frac{\eta^2}{2} \right) - I_{1/4} \left(\frac{\eta^2}{2} \right) - I_{3/4} \left(\frac{\eta^2}{2} \right) \right] \exp \left(\frac{-\eta^2}{2} \right) \quad (11)$$

³ This can be shown in a very simple manner. After integrating over $\bar{\omega}$, we split the remaining integral over K into two parts, one extending from 0 to M/\bar{r} and the other from M/\bar{r} to ∞ . M is a positive number, say 5, large enough so that for $K\bar{r} \geq M$ the asymptotic expression for the Bessel function $J_0(K\bar{r})$ is applicable. It can be shown that for large values of \bar{r} the absolute value of the integral from 0 to M/\bar{r} vanishes more rapidly than does the absolute value of the integral from M/\bar{r} to ∞ . Also, the difference between the latter integral and one containing the asymptotic expression for the Bessel function and extending from 0 to ∞ vanishes more rapidly for large values of \bar{r} .

where $\eta = (\bar{x} - \bar{\omega}_0' \tau)/(2\bar{a}^{1/2})$ and I designates the modified Bessel function of the first kind. The terms involving $I_{-3/4}$ and $I_{1/4}$ are to be understood as even functions, the terms involving $I_{-1/4}$ and $I_{3/4}$ as odd functions of $(\bar{x} - \bar{\omega}_0' \tau)$.

When the point source emits a δ -function pressure pulse, the early response at the detector becomes

$$\frac{\bar{G} |\bar{\omega}_0''|}{4 |a_{10}'| (\pi\bar{r})^{1/2} \bar{a}^{5/4}} \left[\eta 2\Gamma \left(\frac{7}{4} \right) {}_1F_1 \left(-\frac{1}{4}; \frac{3}{2}; -\eta^2 \right) - \Gamma \left(\frac{5}{4} \right) {}_1F_1 \left(-\frac{3}{4}; \frac{1}{2}; -\eta^2 \right) \right]$$

where ${}_1F_1$ designates the Pochhammer-Barnes confluent hypergeometric function [Watson, 1944].

When the point source emits an exponentially decaying pressure pulse, we can readily evaluate the response at the detector by means of numerical convolution of the source excitation with (12).

Especially simple results are obtained for the two-dimensional problem of a line source and a line detector. When the line source emits a Heaviside step pressure pulse, we obtain for the early pressure response

$$\frac{A |\bar{\omega}_0''|}{2 |a_{10}'| \bar{\omega}_0'} \sqrt{\frac{\pi}{\bar{a}}} \exp(-\eta^2)$$

whereas for the δ -function source, we obtain

$$\frac{G' |\bar{\omega}_0''|}{2 |a_{10}'| \bar{a}} \sqrt{\pi} \eta \exp(-\eta^2)$$

Finally, for the line source emitting an exponentially decaying pressure pulse, we have

$$\frac{A |\bar{\omega}_0''|}{2 |a_{10}'| \bar{\omega}_0'} \sqrt{\frac{\pi}{\bar{a}}} \left[\exp(-\eta^2) - \frac{\sqrt{\pi \bar{a}}}{\bar{\theta} \bar{\omega}_0'} \right]$$

$$\cdot \exp \left(\frac{\bar{x} - \bar{\omega}_0' \tau + \frac{\bar{a}}{\bar{\theta} \bar{\omega}_0'}}{\bar{\theta} \bar{\omega}_0'} \right) \operatorname{erfc} \left(\eta + \frac{\sqrt{\pi \bar{a}}}{\bar{\theta} \bar{\omega}_0'} \right)$$

⁴ We have given our solution here in terms of the modified Bessel functions of the first kind since tables of these functions are most readily available. The solution could also have been stated in terms of Pochhammer-Barnes confluent hypergeometric functions, ${}_1F_1(a; b; z)$, or, since b is 1/2 or 3/2, in terms of parabolic cylinder functions.

large contrast and moderate refraction path. We consider only the two-dimensional problem of a line source emitting a Heaviside step pressure pulse. Our analysis proceeds as before, but since \bar{a} is sufficiently small compared with $|\bar{b}|$ we expand \bar{a} in our expansion of the integrand around $\bar{b} = 0$ and substitute third-order terms, using series given in (5) and (6). We obtain as an approximate early refracted pressure response at the detector

$$\frac{A |\bar{\omega}_0''| \pi}{|a_{10}'| \bar{\omega}_0' \bar{b}^{1/3}} \text{Ai} \left(\frac{\bar{x} - \bar{\omega}_0' \tau}{\bar{b}^{1/3}} \right) \quad (16)$$

where Ai designates the Airy function.

It is interesting to note that when the rotational (S) wave velocity of the plate, c_{22} , is appreciably greater than the sound velocity of the fluid, c_1 , there will be a second low frequency 'refracted' arrival at the detector. This arrival is observed by *Osborne and Hart* [1945, 1946] and is associated with an antisymmetric mode. For a source and detector near the plate, this mode effectively starts with an Airy phase (group velocity maximum) of the lowest real antisymmetric mode at a point on the dispersion curve where the horizontal group velocity (but not the horizontal phase velocity) exceeds the velocity of sound in the fluid.

Intermediate case of contrast and refraction path. In this case, $|\bar{a}|^{3/2}$ and $|\bar{b}|$ are of the same order of magnitude. It is not possible to express the integrals (7 to 10) in terms of tabulated functions, so they must therefore be evaluated numerically. If we want to obtain the early pressure response for a source emitting an exponentially decaying pulse, an additional numerical convolution of (8) or (10) with the source excitation function is required.

Discussion. First we point out that $\tau_0 = \bar{x}/\bar{\omega}_0'$ represents the travel time (in reduced units) for a signal which travels from source to detector in the fluid with reduced velocity c_1/c_{22} and on a refraction path along the plate with reduced velocity $\bar{\omega}_0'$. The reader can also readily verify that $(i\bar{\omega}_0' \tau_0)/[a_{10}''(\bar{z} + \bar{d})] = (i\bar{\omega}_0''' \tau_0)/[a_{10}'''(\bar{z} + \bar{d})]$ is the ratio of the total travel time of a signal to its travel time in the fluid medium. Let $\bar{l} = l/H$ be the length (in units of plate thickness) of that section of the above-mentioned refraction path which is along the plate.

A better criterion is $|\bar{b}| \gg |\bar{a}(\bar{x} - \bar{\omega}_0' \tau)|$.

The results for the line source and moderate contrast or long refraction path are particularly simple. Expression 13 for the Heaviside step-function source describes approximately a gaussian curve in time; expression 14 for the δ -function pressure source describes approximately the derivative of a gaussian curve. A very useful approximate description of (14) is given by the statement that the peak pressure amplitudes are

$$P_{\max} \simeq \pm \frac{G(2\pi/e)^{1/2} |\bar{\omega}_0''|}{2 |a_{10}'| [|\bar{\omega}_0''| t_0 - |a_{10}''| (d+z)/c_{22}]} \quad (17)$$

and that these amplitudes occur at

$$t_{\max} \simeq t_0 \mp (\bar{\omega}_0')^{-1} \left(\frac{H}{c_{22}} \right)^{1/2} \cdot \left[|\bar{\omega}_0''| t_0 - |a_{10}''| \frac{(d+z)}{c_{22}} \right]^{1/2} \quad (18)$$

where $t_0 = \tau_0 H/c_{22}$ may be called the 'time of plate-wave arrival.' Note that (17) and (18) are given in ordinary units, and $G = A\theta$ ($A \rightarrow \infty$ and $\theta \rightarrow 0$).

We observe that for the Heaviside step-function source, which also describes approximately the case $\theta \gg H/c_{22}\bar{\omega}_0'$, the peak of the gaussian signal travels along the plate-fluid interface with the longitudinal plate wave velocity $2c_{22}(1 - c_{22}^2/c_1^2)^{1/2}$; for the δ -function source, which also describes approximately the case $\theta \ll H/c_{22}\bar{\omega}_0'$, the zero in the response travels with the plate wave velocity along the interface. For intermediate values of θ (see expression 15) the zero in the response is shifted to later times. Evidently the formulas lend themselves readily to an approximate ray interpretation.

Next we note that for the δ -function source, the peak-to-peak amplitude of the refraction signal varies as l^{-1} and is independent of H . The width of the signal, as measured, for example, from the positive to the negative peak, is proportional to $l^{1/2}$ and also to $H^{1/2}$. For the Heaviside step-function source, the peak pressure amplitude of the refraction signal varies as $l^{-1/2}$ and is proportional to $H^{1/2}$. The width of the signal, as measured, for example, between the inflection points of the gaussian curve, is proportional to $l^{1/2}$ and also to $H^{1/2}$.

For the case of the Heaviside step-function

line source and large contrast and moderate refraction path, the pressure response is given by (16). We see that the peak amplitude is proportional to $l^{-1/3}$ and to $H^{1/3}$. The period of the response, as measured, for example, between peaks or zeros, is proportional to $l^{1/3}$ and to $H^{2/3}$.

Finally, we consider the result for the step-function point source and moderate contrast or long refraction path (expression 11). The peak of the refracted pressure signal arrives approximately at the same time as that for the corresponding two-dimensional problem, but its amplitude varies as $r^{-1/2} l^{-3/4}$ and is proportional to $H^{1/4}$. For the δ -function point pressure source (expression 12), the results are not so clearly seen from the formula. The zero arrives approximately at the same time as for the corresponding two-dimensional problem. The peak-to-peak amplitude appears to vary approximately as

$$\frac{\bar{G} |\bar{\omega}_0''| \exp(-\eta^2)}{4 |a_{10}'| (\pi \bar{r})^{1/2} \bar{a}^{5/4}} \cdot \left\{ \eta \left[\frac{2\bar{b}\Gamma(13/4)}{3\bar{a}^{3/2}} {}_1F_1\left(-\frac{7}{4}; \frac{3}{2}; \eta^2\right) - \frac{2Q\Gamma(9/4)}{\bar{a}^{1/2}} {}_1F_1\left(-\frac{3}{4}; \frac{3}{2}; \eta^2\right) - \frac{\Gamma(5/4)\bar{a}^{1/2}}{4\bar{r}} {}_1F_1\left(\frac{1}{4}; \frac{3}{2}; \eta^2\right) \right] + \left[\frac{\bar{b}\Gamma(11/4)}{3\bar{a}^{3/2}} {}_1F_1\left(-\frac{9}{4}; \frac{1}{2}; \eta^2\right) - \frac{Q\Gamma(7/4)}{\bar{a}^{1/2}} {}_1F_1\left(-\frac{5}{4}; \frac{1}{2}; \eta^2\right) - \frac{\Gamma(3/4)\bar{a}^{1/2}}{8\bar{r}} {}_1F_1\left(-\frac{1}{4}; \frac{1}{2}; \eta^2\right) \right] \right\} \quad (17)$$

where

$$Q = \frac{\rho_2 |a_{10}'|}{6\rho_1} + \frac{5a_{10}''}{2 |a_{10}'|} + \frac{\bar{\omega}_0'''}{3 |\bar{\omega}_0''|}$$

$$\frac{|\bar{\omega}_0''| \left[\left(21 - 62 \frac{c_{22}^2}{c_2^2} \right) \frac{c_{22}^2}{c_2^2} (\bar{\omega}_0')^4 - \left(13 - 21 \frac{c_{22}^2}{c_2^2} - 58 \frac{c_{22}^2}{c_2^4} \right) (\bar{\omega}_0')^2 + \left(25 - 54 \frac{c_{22}^2}{c_2^2} \right) \right]}{4\bar{\omega}_0' \left(3 - 4 \frac{c_{22}^2}{c_2^2} \right) \left(1 - 2 \frac{c_{22}^2}{c_2^2} \right)^2}$$

$r^{-1/2} l^{-5/4}$ and also seems to be approximately proportional to $H^{-1/4}$. This last result seems paradoxical at first sight, since the response appears to become infinite for vanishing plate thickness. It must be remembered, however, that for a physically realizable source, θ is finite; for H sufficiently small, therefore, θ must be $\gg H/c_{22}\bar{\omega}_0'$. Consequently, the formula for the step-function pulse, where the pressure response is proportional to $H^{1/4}$, will eventually be applicable.

In the next section we shall illustrate our results by a numerical example. Since our solution is in the form of an asymptotic series, it is necessary to know how good an approximation we may expect from the leading term of such a series. For this purpose we have evaluated the next higher order term in the solution for the case of moderate contrast. For the δ -function pressure point source, we derive as the next term in a series whose leading term is given by (12),

Similarly, for the δ -function pressure source, the term following (14) is given by

$$\frac{G' |\bar{\omega}_0''| \sqrt{\pi}}{2 |a_{10}'| \bar{a}} \exp(-\eta^2) \left[\frac{\bar{b}}{12\bar{a}^{3/2}} (3 - 12\eta^2) + 4\eta^4 - \frac{Q}{2\bar{a}^{1/2}} (1 - 2\eta^2) \right] \quad (18)$$

Comparison of these terms with the leading terms of the asymptotic series allows us to infer criteria for the range of validity of the asymptotic solution (see also Appendix B).

Numerical example. We illustrate the preceding results by a numerical example for the case of moderate contrast. We choose the case of a 0.159-cm-thick Lucite plate submerged in water. Source and detector are 60 cm apart, and each is 0.3 cm from the surface of the plate; the exponential decay time constant of the source is 6 μ sec, and the amplitude factor A is 1 cm. The following numerical values are applicable

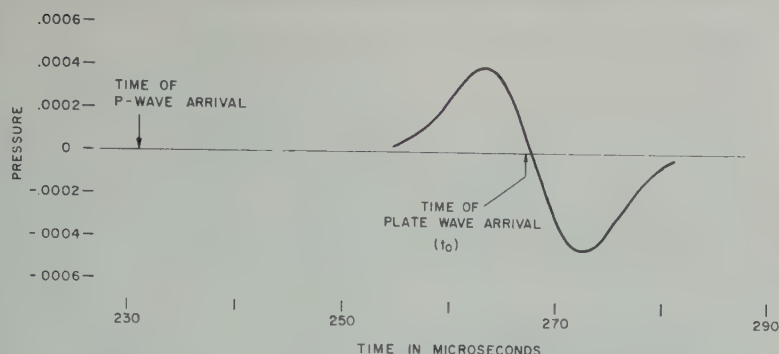


Fig. 2. Pressure signal from a point source refracted through a thin Lucite plate in water. (Note that the peak pressure amplitude of the signal traveling through the water directly from source to detector is 1/60.)

$$\begin{aligned} &= 2650 \text{ m/sec} & \rho_2 &= 1.15 \text{ g/cm}^3 \\ &= 1500 \text{ m/sec} & \rho_1 &= 1 \text{ g/cm}^3 \\ &= 1305 \text{ m/sec} & \bar{\theta} &= 4.925 \end{aligned}$$

$$\bar{r} = 377.3$$

$$\bar{z} + \bar{d} = 3.773$$

$$\bar{A} = 6.29$$

$$\begin{aligned} \bar{\omega}_0' &= 1.741 & |a_{10}'| &= 1.137 \\ |\bar{\omega}_0''| &= 0.1765 & |a_{10}''| &= 0.2045 \\ \bar{\omega}_0''' &= 0.0608 & a_{10}''' &= i 0.0704 \end{aligned}$$

$$Q = -0.661$$

$$\bar{x} = 382.4 \quad \text{and} \quad \tau_0 = 219.2$$

time τ_0 ,

$$\bar{a} = 19.0 \quad \text{and} \quad \bar{b} = -6.54$$

the pressure response at the detector, calculated from the leading term of our asymptotic expansion, is shown in Figure 2.

It is instructive to compare these results with those for the two-dimensional δ -function source, given by (17) and (18). We calculate that

$t_{\max} = 267.1 \pm 4.3 \mu\text{sec}$ and $P_{\max} = \pm 0.0032G \mu\text{sec}^{-1}$. It is interesting to compare these calculations in turn with the case of a δ -function line source and an infinitely thick plate. We find by means of a simple approximate formula, based on equation 4.25 of Strick [1959],⁶ that the peak of the first refracted signal is $0.0036G \mu\text{sec}^{-1}$ and it occurs $7.8 \mu\text{sec}$ after the time of the refracted P -wave arrival.

APPENDIX A

We consider briefly the case of an elastic plate shallowly submerged in a liquid half-space. Let h be the thickness of the shallow liquid layer resting on top of the plate, and let d' and z' be the depth of the source and of the detector, respectively, measured from the free surface of the liquid layer. We set $\bar{h} = h/H$, $\bar{d}' = d'/H$, and $\bar{z}' = z'/H$; \bar{d}' and $\bar{z}' < \bar{h}$.

Using a method of analysis analogous to that employed in the main part of this paper, we find that in the limit as $K \rightarrow 0$, $\bar{\omega} \rightarrow 0$, $\bar{\omega}' \rightarrow 2[1 - (c_{22}^2/c_2^2)]^{1/2}$, $\bar{\omega}'' \rightarrow 0$, and $\bar{\omega}''' \rightarrow -2[1 - (c_{22}^2/c_2^2)]^{1/2} [1 - 2(c_{22}^2/c_2^2)]^2 [1 + (\rho_1 \bar{h}/2\rho_2)]$. Specifically, for the case of a step-function pressure line source in the liquid layer, we find the following leading term for the low frequency early pressure response at the detector

$$\frac{4(2)^{1/3} A\pi(\bar{z}' \cdot \bar{d}') Ai \left\{ \frac{2^{1/3}(\bar{x} - \bar{\omega}_0' \tau)}{\left[\bar{\omega}_0' \tau \left(1 - 2 \frac{c_{22}^2}{c_2^2} \right)^2 \left(1 + \frac{\rho_1 \bar{h}}{2\rho_2} \right) \right]^{1/3}} \right\}}{(\bar{\omega}_0')^{4/3} \left(1 - 2 \frac{c_{22}^2}{c_2^2} \right)^{2/3} \left(1 + \frac{\rho_1 \bar{h}}{2\rho_2} \right)^{1/3} \bar{h} \tau^{1/3}} \quad (\text{A-1})$$

⁶ Note that Strick's P_0 corresponds to $2G$.

Note that this result applies to all cases of contrast, since $\bar{\omega}_0'' = 0$.

APPENDIX B

In connection with a discussion of the quality of asymptotic expressions, it is necessary to emphasize the importance of taking into account singularities which may arise in the vicinity of the point around which we expand the integrand. In many problems concerning the propagation of elastic waves in layered elastic media, the dispersion curves of the characteristic vibrations show large changes in the group velocity for small changes in the wave number. One of the simplest cases in point is the problem of the liquid layer over a liquid bottom. If the acoustic contrast between the liquids in the layer and in the bottom is appreciable, then the horizontal group velocity U , corresponding to the first arrival refracted along the bottom interface (whose frequency and horizontal wave number are at the so-called low-frequency cutoff values $\bar{\omega}_c$ and K_c), changes rapidly with the horizontal wave number K even for the lowest mode. When we consider the response of this system to explosive sound we are led to the evaluation of an integral over the horizontal wave number K by the method of stationary phase (or of steepest descent), and a difficulty arises when we are interested in early refraction arrivals. We find that the derivatives $\bar{\omega}_c''$, $\bar{\omega}_c'''$, etc., become increasingly large, so that our solution is appropriate only for extremely large horizontal ranges. It appears, however, that this difficulty can be avoided by means of a simple change of variable. The large size of the derivatives can be ascribed to the presence of a branch point of the function $\bar{\omega}$ in the K plane. This branch point is of the square-root type and is located at K_b ,

close to the cutoff value K_c . (This branch point must not be confused with a branch point at κ of the period equation considered as a function of the two independent variables $\bar{\omega}$ and K .) We replace K by the new independent variable $\kappa = (K - K_b)^{1/2}$, where K_b is readily evaluated [Rosenbaum, 1960, eq. 53]. The higher derivatives of $\bar{\omega}$ with respect to κ at $\kappa_c = (K_c - K_b)^{1/2}$ are now of moderate size, and the methods of stationary phase or of steepest descent should yield satisfactory answers.

The example given above should emphasize the point that each solution for a specific mode may require careful attention.

Acknowledgments. I wish to thank Dr. A. Ginzburg, Dr. G. W. Postma, Mr. H. Rainbolt, Dr. A. J. Seriff, and Dr. E. Strick of the Shell Development Company for many helpful discussions. I am also indebted to Mr. W. L. Roever for experimental data, which pointed the way to some of the results presented here.

REFERENCES

- Osborne, M. F. M., and S. D. Hart, Transmission, reflection, and guiding of an exponential pulse by a steel plate in water, 1, Theory, *J. Acoust. Soc. Am.*, **17**, 1-18, 1945.
- Osborne, M. F. M., and S. D. Hart, Transmission, reflection, and guiding of an exponential pulse by a steel plate in water, 2, Experiment, *Acoust. Soc. Am.*, **18**, 170-184, 1946.
- Rosenbaum, J. H., The long-time response of a layered elastic medium to explosive sound, *Geophys. Research*, **65**, 1577-1613, 1960.
- Strick, E., Propagation of elastic wave motion from an impulsive source along a fluid/solid interface, 2, Theoretical pressure response, *Phil. Trans. Roy. Soc. London A*, **251**, 465-488, 1955.
- Watson, G. N., *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, p. 100, 1944.

(Manuscript received July 13, 1961; revised August 24, 1961.)

Energy Requirements in Terrestrial Expansion¹

MELVIN A. COOK AND A. J. EARDLEY

*Institute of Metals and Explosives Research
College of Mines and Mineral Industries, University of Utah
Salt Lake City, Utah*

Abstract. Terrestrial expansion is investigated in terms of its energy requirements. Uniform expansion of the radius by 20 per cent would increase the gravitational energy by almost the amount required to dissociate all chemical bonds of molecules constituting the earth. Chemical changes, therefore, probably cannot be considered responsible for such expansion. The increase of gravitational energy for a uniform expansion of 20 per cent would be at least 10^8 times greater than the radiogenic heat produced in the earth since the Paleozoic (225 m.y.), according to generally accepted estimates of the abundances of radioactive elements. Expansion of this amount due to temperature rise is therefore improbable. Phase changes are likewise shown to be inadequate for appreciable expansion in the face of gravitational energy requirements. Upward and downward propagation of phase boundaries, should they be induced by ice cap loading and unloading, and consequent isostatic adjustments are investigated and concluded to be improbable.

Interest in the concept of appreciable expansion of the earth is currently high [Carey, 1958; Eyring, 1959; Eged, 1956] because the idea presents a plausible mechanism for the separation of continents and the growth of the ocean basins. To account for a proposed separation of original single continent to fragments in the present continental positions Carey proposed an increase in the earth's surface area of 45 per cent since the Paleozoic (225 m.y.). This would be a 20 per cent increase in the radius at a rate of about 0.5 cm/year. Eged assumed an expansion rate of 0.04 to 0.08 cm/year, which is about one-tenth of Carey's estimated rate. Paleomagnetic evidence, marshalled by Cox and Doell [1961], indicated that the earth's magnetic field now is about the same as in Permian time. The widely recognized limits of error in their analysis, however, which might admit of an expansion

Shortly after completion of the manuscript of this paper an article by A. E. Beck entitled 'Energy Requirements of an Expanding Earth,' appeared in the *Journal of Geophysical Research* (vol. 66, No. 5, 1485-1498, 1961). Beck's treatment of the gravitational energy requirements in expansion is considerably more refined and detailed than that of the writers, but his results and conclusions are in very close agreement. In the present article we discuss, in addition, limitations imposed on expansion by the observed terrestrial temperatures and on rapid isostatic adjustments of the crust by energy sources, heat conduction, and the pressure-temperature profile.

by as much as Eged postulated, but definitely not by the amount that Carey considered necessary.

Gravitational energy. The rate of increase of gravitational energy ($E = -fGM^2/R$) due to expansion of the earth is given by the differential equation

$$\frac{dE}{dt} = \frac{fGM^2}{R^2} \frac{dR}{dt} - \frac{GM^2}{R} \frac{df}{dt} \quad (1)$$

where dE/dt is the rate of gravitational energy increase of the earth due to expansion (neglecting accretion) taking the earth's mass M as a constant, R is the average radius of the earth, and G is the gravitational constant ($6.67 \cdot 10^{-8}$ in cgs units). The factor f expresses the influence of the density distribution on energy which must be taken into consideration in any model of earth expansion. For example, if the density ρ were assumed to be the same throughout the earth, one would have $f = 0.60$ and $df/dt = 0$. The widely accepted density-radius contour based on the data of seismologists requires $f \approx 0.68$ as computed, for example, from the density data summarized by Kuiper [1954]. That the seismologist's model of the earth, i.e., their $\rho(R)$ relations, are likely to be essentially correct was strikingly demonstrated in recent studies of the Chilean earthquakes wherein excellent agreement was found between predictions and observations of the periods of

the various modes of free oscillations of the earth [Benioff, Press, and Smith, 1961; Ness, Harrison, and Slichter, 1961; Alsop, Sutton, and Ewing, 1961].

A change in density over the whole earth assumed to maintain the present density distribution in time would require $df/dt = 0$. Then (1) would become

$$\frac{dE}{dt} = 0.68 \frac{GM^2}{R^2} \frac{dR}{dt} = 4 \times 10^{30} \frac{dR}{dt} \frac{\text{ergs}}{\text{sec}} \quad (1a)$$

Assuming Carey's suggested expansion of 0.5 cm/year and Egyed's postulated expansion of 0.06 cm/year, one requires in an assumed uniform-in-time terrestrial expansion about 5×10^{22} ergs/sec and about 6×10^{21} ergs/sec, respectively. In 225 m.y. this would amount to approximately 5×10^{28} ergs in Carey's model and 6×10^{27} ergs according to the rate postulated by Egyed.

Chemical bonding energy. If an average atomic weight of 30 and an average chemical bond strength of 10^{-11} ergs/bond (6.3 ev) are assumed for the material constituting the earth, the total chemical energy of all substances of the earth combined would amount to about 10^{30} ergs. Hence, if chemical energy were to have caused the expansion postulated by Carey in a uniform terrestrial expansion model, nearly all the chemical bonding energy of the present earth would have been required to produce it. This would then require the impossible condition that all molecules would have been effectively dissociated in the Paleozoic in order, upon association, to supply the necessary energy for expansion by chemical means. Even the smaller radial expansion proposed by Egyed, which would have been too small by a factor of 10 to explain the present continental distribution by simple expansion, would seem very unlikely. Note, for example, that if the whole earth were composed of TNT, its explosion would cause an expansion of only 250 km or 4 per cent, assuming f to remain essentially unchanged. Thus, the only way to expand the earth *uniformly*, i.e., without invoking either accretion or phase changes, would be by uniform radioactive heating.

Radioactive decay energy. Uniform radioactive heating of the earth would require heat generation at an average rate of about 2×10^{28} (Egyed) to 2×10^{30} (Carey) ergs/year. The heat loss conventionally attributed to radio-

activity in the earth is about 10^{28} ergs/year, which is only 0.05 to 0.005 the total rate of heat generation required to expand the earth at the rates suggested by these expansion models. The 10^{28} ergs/year has been shown to agree essentially with that generated radiogenically in the earth's crust, and it is generally agreed that this level of radioactivity is much lower in the mantle and core of the earth than in the crust. For example, the total rate of heat generated radiogenically in the mantle is estimated by Kuiper to be only about 10 per cent of that in the crust. Even if radioactive elements (U, Th, etc.) were considered to be uniformly distributed throughout the earth at the same concentration as in granite, and no heat were lost from the earth, one could account for a $\Delta R/\Delta t$ of only 0.01 cm/year. This is too small for Carey's model but would account for Egyed's proposed expansion.

Terrestrial expansion by radioactivity would produce expansion by release of heat, i.e., thermal expansion. But the coefficient of expansion is so low as to require absurd present terrestrial temperatures if expansion of the suggested magnitude has occurred. For example, even at an average bulk expansion coefficient of $2 \times 10^{-5}/^\circ\text{C}$, which is unrealistically high under the high pressures involved, an average temperature increase of about 4000°C is required to expand the earth 7.5 per cent and about $40,000^\circ\text{C}$ to expand it 75 per cent in volume. The apparent temperature profile of the earth does not, therefore, indicate appreciable thermal expansion because the average temperature of the earth is probably even yet less than 4000°C [Kuiper, 1954]. Furthermore, if the earth had been heated radiogenically by 4000° to $40,000^\circ\text{C}$ since the Permian, under the uniformity concept it should have heated about 10^5 to 10^6 degrees since Early Precambrian. Furthermore, the rate of heat loss from the surface would be many times higher than it is. Clearly radioactivity cannot have caused the postulated expansion.

Phase changes and expansion. Writing equation 1 in terms of the mass M and present radius R of the earth, one has

$$\begin{aligned} \frac{dE}{dt} &= 2.5 \times 10^{30} \frac{d}{dt} (\ln R) \\ &\quad - 3 \times 10^{30} \frac{df}{dt} \quad (\text{in cgs units}) \end{aligned} \quad (1)$$

by the value of f computed from the accepted terrestrial density distribution is actually quite close to the value for a uniform sphere under uniform pressures in a relatively cool earth. The difference ($\Delta f = 0.08$) between the computed and that of a constant density sphere is, in other words, practically entirely accounted for by the influence of pressure. In the last 5×10^9 years f must thus have increased much less than 0.08, and therefore $\Delta f/\Delta t$ must be very much less than $0.08/5 \times 10^9 \doteq 10^{-11} \text{y}^{-1}$. Therefore, apparently no model of expansion whatever can account for a sufficiently large last term in (1) to permit expansion with appreciably less energy than required by (1a) because df/dt cannot have been appreciable for the earth.

In estimating the influence of pressure alone one may make use of the theoretical equation of state [Cook, 1956, 1959],

$$\beta/\beta_0 = (\rho_0/\rho)^a = (1 + a\beta_0 p)^{-1} \quad (2)$$

which has the same form as the empirical Tait [1898] equation of state. Here β is the compressibility, ρ the density, p the pressure, a a constant, and subscript 0 refers to values at atmospheric pressure. Equation 2 has been used by studies in this laboratory to fit well with extensive high shock pressure data of Rice, McQueen, and Walsh [1958] using empirical values of a falling in the range of 2.0 to 6.0. This equation fitted to experimental data extending up to about 0.5 megabars gives densities at the center of the earth in good agreement with the densities estimated by seismologists [Weiper, 1954].

Clearly one has no leeway for manipulation of f once the influence of pressure on density is taken into account. But in spite of the fundamental difficulty imposed by the seismological evidence that $\Delta f/\Delta t$ must be quite negligible let us consider the possibility of appreciable expansion by means of phase changes. Appreciable expansion due to phase changes must necessarily cause f to change appreciably in time or the expansion could not actually occur for reasons involving energy. In order, therefore, to cause expansion of the earth's radius R at a uniform rate of about 0.5 cm/year or to give a total discontinuous expansion of 20 per cent (or more than 10^3 km) in R one really requires, because of energy limitations, that the second term on the right-hand side of (1) must be al-

most as large as the first term. That is, as shown by (1), dE/dt can be realistically small only if

$$\frac{df}{dt} \approx \frac{d}{dt} (\ln R)$$

An increase in time in f means, of course, a rapid rate of increase in the earth's density gradient measured inward from the surface of the earth. This means, furthermore, that the expansion rate must be greatest toward the surface and least toward the core of the earth in a model that permits

$$R^2 dE/dt \ll fGM^2 dR/dt$$

The only apparent means for df/dt to be appreciably positive would be for postulated phase changes to occur near the surface of the earth. Jeffreys [1959] has summarized the seismic evidence for an abrupt discontinuity in the mantle at about 480-km depth and others [Mohorovičić, 1925; Bernal, 1936; Dahm, 1936; Birch, 1951] have discussed the evidence for other phase changes above 1000-km depth. Hoffman, Berg, and Cook [1961] observed seismic reflections or density discontinuities from large commercial explosions at depths of 190, 210, 520, 555, and 910 km, closely confirming earlier observations connected with atomic explosions [Carder and Bailey, 1958]. That phase changes exist in the outer layers of the earth seems most likely; the only question is whether these phase boundaries have always existed at the same depths at which they are now found, or whether they have been moving steadily downward in time. They can occur only at certain pressure and temperature conditions, of course; hence one needs to consider whether the possible $p(T)$ contour of the earth will permit appreciable movements of phase boundaries in time.

First, to see whether phase transitions can expand the earth appreciably, let us make use of the following somewhat exaggerated assumptions: let us assume (1) that all observed discontinuities are due to phase changes, (2) that at the end of the Paleozoic all discontinuities were close to the surface and that since then they have propagated steadily or discontinuously to their present positions, and (3) that the density changes, lumped over a postulated single discontinuity, have averaged as large as 0.7 g/cm^3 in the upper 900 km of the mantle, as

suggested by *Heiskanen and Vening Meinesz* [1958]. The total volume increase due to all possible phase changes in the region where they can contribute relatively low energy expansion would be given by

$$\Delta V \leq \Delta V_{\infty} + \Delta V_{\infty} + \Delta V_x \quad (3)$$

where ΔV is the total volume increase; ΔV_{∞} is the maximum possible volume increase due to an assumed change in depth of the Mohorovicic discontinuity from near zero to its present 5-km depth below the ocean; ΔV_{∞} is the maximum possible volume increase from the postulated similar phase change assumed to have occurred from the surface to about 30-km depth under the continents; and ΔV_x is the maximum possible increase in volume due to postulated phase changes assumed to be now at a depth of 900 km and to have moved down uniformly or discontinuously during the period in question all the way from the surface of the earth. We thus compute as the most optimistic upper limit of expansion by means of phase changes

$$\Delta V \leq 1.6 \times 10^{23} + 8 \times 10^{23} + 8 \times 10^{25} = 8.1 \times 10^{25} \text{ cm}^3$$

It is realized, of course, that many of these supposed phase changes are not phase changes at all, but rather discontinuities due to chemical differences, and that none of them could have propagated as far as indicated. Since the present volume of the earth is about $1.1 \times 10^{27} \text{ cm}^3$ all possible or vividly imagined phase changes would have expanded the earth less than 8 per cent in volume or 2.5 per cent in radius, not merely since the Permian but since the Early Precambrian.

Note that more than 90 per cent of the possible expansion indicated by (2) is associated with the assumed discontinuity actually centered, not at 900-km depth, but at about 480-km depth, or possibly extending uniformly or non-uniformly to about 900-km depth as suggested by *Heiskanen and Vening Meinesz*. Therefore, ΔV_x would really be no more than $4 \times 10^{25} \text{ cm}^3$ even in a most optimistic estimate. Hence the minimum energy associated with phase changes would be

$$\Delta E = \int p \, dV = \bar{p} \, \Delta V \doteq 10^{37} \text{ ergs}$$

Thus even an upper-mantle expansion would require an excessive energy, and expansion due to phase changes at still greater depth would require a still greater gravitational energy increase. It is thus clear that the value of $d\bar{f}/dt$ cannot exceed $0.9 \, d(\ln R)/dt$ and that phase changes therefore cannot account for the expansion required in these earth expansion models.

Isostatic adjustment by means of phase changes. Other fundamental arguments against propagation of phase boundaries inside the earth pertain on the one hand to the impossibility of adiabatic transitions at random depths and on the other to limitations in heat transfer. The specific enthalpy ϵ of a phase change is given by

$$\epsilon = \epsilon_0 + \bar{p} \, \Delta v_x \quad (4)$$

where ϵ_0 is the specific enthalpy at zero pressure, \bar{p} is the average pressure (in the present case corresponding to $\bar{h} \doteq 500 \text{ km}$, or about 165 kb) and Δv_x is the volume change per gram. For an average Δv_x of $0.05 \text{ cm}^3/\text{g}$ corresponding to the above computations of upper-mantle expansion the last term in (4) would amount to about 200 cal/g.

Now, in order for the boundary of a supposed phase change to propagate at an appreciable rate, the pressure-temperature curve of the phase transition would need to match closely the $p(T)$ contour of the earth. This is true because the heat generation and, as seen from the observed heat flux from the earth, the heat conduction itself would be too small to permit an appreciable time rate of change of the $p(T)$ curve of the earth in the region of the postulated phase change. But dp/dT for the mantle is only about 700 ergs/cm³ deg K. This requires, according to the Clapeyron equation,

$$(dp/dT) = (\epsilon/T \, \Delta v) = (T \, \Delta s)/(T \, \Delta v) = \Delta S/\Delta V \doteq 700 \text{ ergs/cm}^3 \text{ deg K}$$

Thus, for this to apply, ϵ would have to be negligibly small, since $\bar{p} \Delta v = 8 \times 10^9 \text{ ergs/g}$. One thus requires for the entropy change ΔS only about 35 ergs/cm³ deg K; $T \Delta s \doteq 7 \times 10^9 \text{ ergs/g}$, and therefore $\epsilon_0 \doteq -\bar{p} \Delta v_x = -8 \times 10^9 \text{ ergs/g}$. The condition $\epsilon_0 = -\bar{p} \Delta v_x$ is really a very restricted one, permitting a phase change to occur only at a fixed depth, namely that depth at which the pressure is $-\epsilon_0/\Delta v_x$. Since

sure changes rapidly with depth, it is clear that a long-range propagation of phase boundary (transition region) deep in the earth at any significant rate, or without exceedingly high rates of heat transfer, is really quite impossible. In other words, one requires a significant change in the absolute temperature at a given depth in order for the phase boundary to propagate. Even in the crust, where radioactivity is considered to be greatest, the temperature increase should not exceed about 10^{-6} deg C/year if no heat were being lost; actually, heat is being lost at a rate about equal to its rate of (radioactive) generation [Kuiper, 1954].

Let us now consider the Heiskanen and Vening Meinesz suggestion of a rapid upward and downward propagation of the postulated phase changes at 200- to 900-km depth which was used to explain observed upward and downward movements of the crust in Fennoscandia and other isostatically adjusting areas under differential loading at the surface, i.e., by ice accumulation and sudden denudation. To explain the observed uplift of nearly 600 m in Fennoscandia, if this area was apparently suddenly denuded of ice at a time which studies suggest was about 10,000 years ago, Heiskanen and Vening Meinesz suggested an expansion in the region thought to be 200- to 900-km deep. The required total upward movement of an effective or average boundary excluding other 'slower' mechanisms, to account entirely for this process (since $\rho \approx 0.05$ cm³/g) should therefore be $6 \times 10^4 / 0.05 = 3.7 \times 10^6$ cm, the rate of rise ranging from about 150 cm/year at first to nearly 6 cm/year at present (using their data). They also required $\Delta s = 0.1$ cal/deg g and $T\Delta s = 201$ cal/g. The total energy involved in this expansion would then be 2.5×10^8 cal/cm².

The depression of 600 m by the ice at the beginning of this uplift, just before sudden denudation, required about 2-km depth of ice because the density of ice is only about one third that of the mantle. The potential energy of the ice load was thus $mgh_0 \approx 2.10^{13}$ ergs/cm² or 0.5×10^8 cal/cm². Therefore, the mechanism of Heiskanen and Vening Meinesz would amplify energy at least 500-fold! That is, by loading the surface of the earth with an ice load of given potential energy relative to the ocean level from which the ice was taken, about 500 times as much heat would be generated at a downward-

moving phase boundary at 500-km depth. If this is not actually impossible, at least it emphasizes the requirement either that the enthalpy of the process would have to be much smaller than the postulated 200 cal/g as explained by equation 4 with $|\epsilon| \ll |\epsilon_0|$, or that heat conduction through a given horizontal plane at a given depth would have amounted in the last 10^4 years to 2.5×10^4 cal/year, a value about a thousand times the heat flow rate estimated from other evidence [Kuiper, 1954]. Bullard showed, for example, that the radiogenic heat from the crust should contribute 9×10^{-7} cal/cm² sec or 36 cal/cm² year, which accounts for practically the entire heat flux at the earth's surface. Since the main heat source is actually above the discontinuity in question, the heat flow through horizontal layers below the crust would be much smaller; the radiogenic heat source in the crust actually would contribute practically all of the heat flux. But the predicted temperature gradient at 500-km depth is only about 4.5°C/km, compared with 30°K/km in the crust. It seems very likely, therefore, that isostatic adjustments cannot be associated with phase changes occurring deep in the earth, but rather with circulations in the mantle, perhaps immediately under the crust.

Conclusion. No other means of expansion besides thermal expansion and phase changes have been suggested. Consideration of these models leads to the conclusion that the earth could not have expanded to the extent proposed by some of the advocates of continental separation.

REFERENCES

- Alsop, L. E., G. H. Sutton, and M. Ewing, Free oscillations of the earth observed on strain and pendulum seismographs, *J. Geophys. Research*, **66**, 631, 1961.
- Benioff, H., F. Press, and S. Smith, Excitation of free oscillations by earthquakes, *J. Geophys. Research*, **66**, 605, 1961.
- Bernal, J. S., Geophysical discussions, *Observatory*, **59**, 268, 1936.
- Birch, F., Remarks on the structure of the mantle and its bearing on the possibility of convection currents, *Trans. Am. Geophys. Union*, **32**, 533, 1951.
- Carder, D. S., and L. F. Bailey, Seismic wave travel times from nuclear explosions, *Bull. Seism. Soc. Am.*, **48**, 377, 1958.
- Carey, S. W., *Continental Drift, Symposium, University of Tasmania*, 177, 1958.

- Cook, M. A., Mechanism of cratering in ultra-high velocity impact, *J. Appl. Phys.*, **30**, 729, 1959; Compressibilities of solids and the influence of inert additives on detonation velocity in solid explosives, *Discussions Faraday Soc.*, No. 22, 207-208, 1956.
- Cox, A., and R. R. Doell, Palaeomagnetic evidence relevant to a change in the earth's radius, *Nature*, **189**, 45, 1961.
- Dahm, C. G., Velocities of *P* and *S* waves from the observed travel time of the Long Beach earthquake, *Bull. Seism. Soc. Am.*, **26**, 1936.
- Egyed, L., Determination of changes in the dimensions of the earth from palaeogeographical data, *Nature*, **178**, 534, 1956.
- Heezen, B. C., Preprints, edited by M. Sears, *International Oceanog. Cong.*, **26**, 1959.
- Heiskanen, W. A., and F. A. Vening Meinesz, *The Earth and Its Gravity Field*, McGraw-Hill Book Co., New York, 1958.
- Hoffman, J. P., J. W. Berg, and K. L. Cook, Discontinuities in the earth's upper mantle as indicated by reflected seismic energy, *Bull. Seism. Soc. Am.*, **51**, 17, 1961.
- Jeffreys, H., *The Earth*, Cambridge University Press, 4th ed., 1959.
- Kuiper, G. P., *The Earth as a Planet*, University of Chicago Press, 1954.
- Mohorovičić, S., see P. Byerly, The Montserrat earthquake of June 28, 1925, G.M.C.T., *Bull. Seism. Soc. Am.*, **16**, 209, 1926.
- Ness, N. F., J. C. Harrison, and L. B. Slichter, Observations of free oscillations of the earth, *J. Geophys. Research*, **66**, 621, 1961.
- Rice, M. H., R. G. McQueen, and J. M. Walsh, Compression of solids in strong shocks, *Shock State Physics*, **6**, Academic Press, N. Y., 1961.
- Tait, P. G., *Scientific Papers*, vol. 2, Cambridge University Press, 1898.

(Manuscript received June 26, 1961; revised August 7, 1961.)

Permeability Measurements of Rock Salt

EARNEST F. GLOYNA AND TOM D. REYNOLDS

*Civil Engineering Department, Sanitary Engineering Research Laboratories
University of Texas, Austin*

Abstract. A method is described for measuring the permeability of rock salt. Different specimens of salt were studied, and the technique was checked by use of a solid crystal of sodium chloride. Both reactive and nonreactive fluids were used in collecting permeability data. Permeability tests were run on rock salt from both dome and bedded formations. The dome salt was found to have a very low permeability, and the bedded salt was found in general to be impermeable. Where there was some measurable permeability under the conditions of the test, it was found that the permeating flow occurred through cracks or fractures in the salt and not through the crystals themselves. These fractures were apparently caused by the relaxation of stress which occurs when underground samples are removed from a compressed formation. The dome salt was found to have more fractures than the bedded salt, owing probably to its lower elastic strength. Such surface fractures, found in both types of salt, will be localized in the vicinity of a mine and consequently will not extend throughout a formation.

Introduction. The troublesome wastes originating from reactor-fuel reprocessing plants are usually stored in artificial containers, and it is highly desirable to dispose of these wastes by the permanent storage scheme. The use of rock storage must be considered a temporary measure because the lifetime of a tank is short compared with that of the radioactive portion of the wastes. In all respects, the problems involved in the disposal of these wastes into any environment are quite complex. Thick shielding and remote handling schemes are required, and in most cases the wastes are corrosive. Heat generation from radioactive isotope disintegration may require refrigeration or at least dilution to the point at which rises in temperature do not interfere with the effectiveness of the containment system.

The Committee on Waste Disposal, Division of Earth Sciences, National Academy of Sciences, has recommended that storage of radioactive wastes in salt cavities represents the most promising permanent solution [Natl. Acad. Sci. U.S., 1957]. This committee of leading scientists in interrelated fields such as sanitary engineering, chemistry, physics, geology, geophysics, radiography, and economics concluded that concentrating the waste and then isolating it from the environment was more practical than diluting it to an allowable level of radioactivity. It was recognized that for long-range planning the dilution of high-level radioactive wastes is

utterly impossible. However, before concentrated radioactive wastes can be stored in salt formations, it is necessary to know more about the permeability of rock salt, the build-up of the temperature in a cavity as a result of γ -ray interaction with the salt and waste, and the specific structural characteristics of the plastic and elastic salt masses. The structural problems due to temperature changes have been considered in other papers and reports [Schechter and Gloyna, 1959; Serata and Gloyna, 1960; Heroy, 1956].

The permeability of rock salt is of considerable importance because the rock salt must contain the radioactive nuclides so as to prevent the contamination of valuable resources by them. Salt samples removed from a salt mine for testing are not, however, typical of the salt in an undisturbed location. When a sample is removed, a certain amount of relaxation occurs, and cracks usually appear. It is for this reason that overburden loads must be placed on test samples if representative permeability values are to be obtained.

Test apparatus. Each test stand consisted of a permeability test cell, an assembly for applying the confining load, a unit for forcing the fluid through the specimen, and a flow-measuring device. Three test cells and stands were used. The entire test equipment was in a room where the temperature was maintained at $79^{\circ} \pm 2^{\circ}\text{F}$. A schematic diagram of the

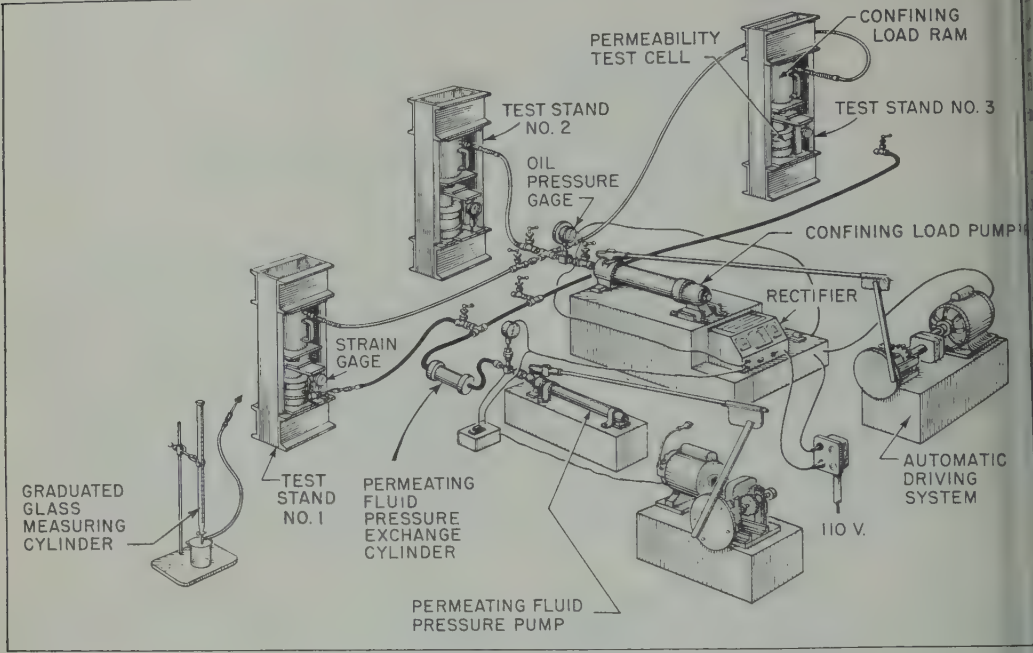


Fig. 1. Assembly of comprehensive test device.

comprehensive test device used during the liquid test is shown in Figure 1.

The test-cell assembly consisted of a specimen holder, a piston, a rectangular compressor, and an encased salt specimen. A typical unit is

shown in Figure 2. The holder and major components were made of special high-grade stainless steel. This particular design was chosen because it provided flexibility. The device was designed so that external heat could be applied, and a fluid used in lieu of the plastic.

A similar test assembly was used to study the flow of helium through the salt specimens. When gas was used as the permeating fluid, appropriate manometers and volume measuring devices were used.

Test procedures. The tests were performed on salt taken from both dome and bedded deposits as well as on a synthetic crystal of sodium chloride. The dome specimens were obtained from the Morton Salt Company's mine at Grand Saline, Texas. The salt was mined as a single block, about 3 ft³ in size, from a point about 700 feet below the earth's surface. The block of rock salt was removed with extreme care; power saws and drills were used and excessive cracking was prevented. The bedded salt was obtained from the Carey Salt Company mine at Hutchinson, Kansas. The block was moved at a point in the mine which is 645 feet below the earth's surface. The block had a volume of 3.5 ft³ and was mined in a similar

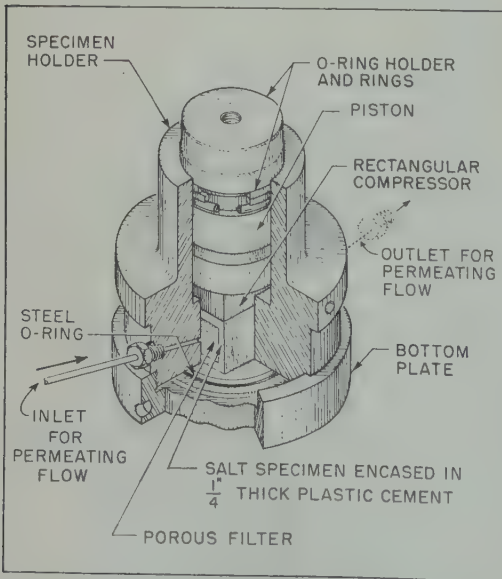


Fig. 2. Permeability test cell assembly.

ner to that for the Grand Saline specimen. The blocks were sprayed with a plastic film to prevent moisture condensation. The pure crystal NaCl was obtained from the Harshaw Chemical Company.

The permeating fluids used during various tests were kerosene, mineral oil, saturated brine solution (NaCl), saturated brine and calcium sulfate solution (NaCl + CaSO₄), saturated brine and aluminum nitrate [NaCl + Al(NO₃)₃], and helium gas.

Both the dome and bedded specimens of salt were prepared in the same manner, but the dome salt required more care in handling because it was more fragile. This was probably due to the larger crystal structure and the higher degree of purity. Crude test specimens measuring approximately 1¾ by 1¾ by 2½ inches were cut from the large block of salt with a handsaw. The crude specimens were then finished with sand paper to the desired size of 1¾ by 1¾ by 11⅛₁₆ inch. The finished specimens were air-cleaned and stored in a desiccator. Typical finished and encased samples are shown in Figures 3 and 4.

The placement of the specimen in the test cell must be done with care. The first step in preparing a finished specimen for a test was to fit porous ceramic filters over the two smaller faces of the specimen, the second step was the placement of the specimen with its ends in the confining chamber of the test cell, and the third step was the actual encasement of the specimen on four sides by the plastic cement. Tests have shown that at pressures above 300 lb/in² this plastic performs in the same manner as a true fluid [Serata and Gloyna, 1959]. The plastic resin cement, Hysol, was

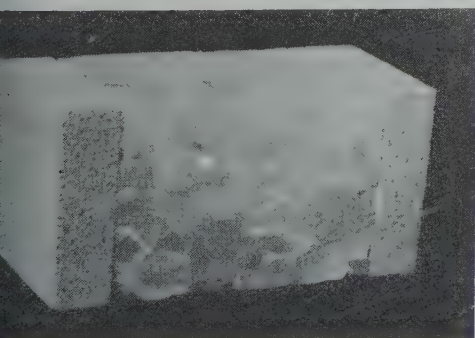


Fig. 3. Finished specimen with end filters.

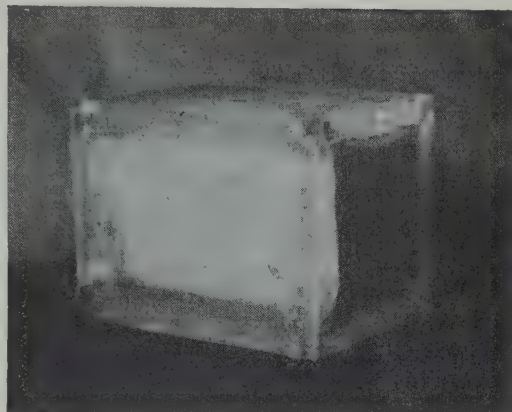


Fig. 4. Encased specimen after completion of test.

mixed in a proportion (8:1) with its hardening agent such that the hardened cement produced a stress-strain curve approximately equal to that of the salt when subjected to structural tests.

The final step in the preparation of a specimen for testing consisted in putting the cell assembly together. The assembled cell was placed in the test stand, and the confining load was applied.

The tests were generally conducted with a single confining pressure of either 500, 1000, 1500, 2000, or 2500 lb/in². Once the desired confining pressure was applied, the loading system was placed on automatic operation. If only liquid permeability tests were to be made, the specimens were placed under a permeating fluid pressure of 150 to 200 lb/in² and maintained at this pressure until the permeability tests were begun. If, however, a specimen was to be tested with gas, no liquid was applied, and the specimen was kept dry.

The creep, or vertical deformation, was recorded daily. When the creep became negligible, the permeability tests were begun. Generally, permeability measurements were started between 7 and 14 days after initial loading.

Daily permeability tests were made in accordance with the procedures recommended by the American Petroleum Institute [1952]. The tests involving liquids were conducted at a permeating-fluid pressure of 100, 200, 300, 400, and 500 lb/in². The permeating-fluid pressure was always kept below a value of one-third of the confining pressure. Differential pressures

TABLE 1. Permeabilities of Hutchinson Salt

Test	Confining Pressure P_c , lb/in ²	Average Fluid Pressure p_f or p_m , lb/in ²	Mean Stress σ_m , lb/in ²	Permeability	Max. Permeability, cm ²	Min. Permeability, cm ²	Test Period, days	Days of Flow	Days of Flow
1H	2450	150	2300	k_K	2.3×10^{-13}	5.0×10^{-15}	23	23	0
2H	1100	100	1000		0	0	7	0	7
3H	1100	100	1000	k_B	7.2×10^{-14}	0	9	3	6
3H	2150	150	2000	k_B	1.5×10^{-14}	0	11	6	5
4H	1510	10	1500	k_G	0	0	2	0	2
5H	510	10	500	k_G	0	0	2	0	2
5H	550	50	500	k_K	0	0	2	0	2

which were one-tenth of the liquid pressures were used in the gas tests.

The effective porosity was determined by the gas expansion method [Washburn and Bunting, 1922] in which a mercury pump porosimeter was used.

The permeability values for each test are designated by the letter k with a subscript to denote the fluid used:

k_G = permeability to helium gas.

k_K = permeability to kerosene.

k_O = permeability to mineral oil.

k_B = permeability to brine (NaCl).

k_C = permeability to calcium sulfate and brine ($\text{CaSO}_4 + \text{NaCl}$).

k_S = permeability to synthetic waste of aluminum nitrate and brine ($\text{Al}(\text{NO}_3)_3 + \text{NaCl}$).

The flow through a specimen was found to occur through the interconnected fractures. That no flow occurs through the crystals was proved by testing a single crystal obtained from the Harshaw Chemical Company. The test produced no flow, and it was therefore assumed that the techniques involved in the tests were satisfactory. If flow occurs in a specimen of rock salt, it must occur through the intercrystalline planes or fractures.

Special procedures were used on one specimen to determine whether the plastic encasement cement was causing a prestress condition and thereby changing the apparent conditions. To test this possibility, the sides of a specimen were coated with a thin layer of Duco cement, and it was then mounted in the test cylinder. The specimen was tested with helium gas both

before and after the Hysol mixture was poured into the void between the specimen and cell assembly. The permeabilities were found to be approximately equal. Since the permeability compared favorably, it was assumed that the Hysol was not affecting the specimen significantly.

Comparison of rock salts. The Hutchinson salt had a lower permeability than the Grand Saline salt. In both cases the permeabilities were small and subject to consolidation of the specimens. Both the amount of overburden load and the duration of the load are important aspects when determining the *in situ* permeability of salt specimens. The fact that some salt from the bedded formations produced no flow during the permeability tests resulted from structural consolidation, which is more readily attained with salt of less purity. It should be noted that the permeability of both types of salt may be affected by interacting agencies. Under the conditions of these tests, the bedded salt appeared either to reach a higher degree of consolidation sooner or to have impurities which restricted the flow of permeating fluid. If the difference in apparent permeability is due to consolidation rates used in these tests, it is of no importance. However, if the difference is due to a chemical interaction, considerable importance must be attached to it.

Seven tests were run on Hutchinson salt at various confining pressures. Helium gas, pure brine solutions, and kerosene were used as fluids. Values are shown in Table 1. The permeability values were less consistent than values similarly obtained using the salt from the Grand Saline mine. Of the seven tests performed v

hinson salt, four showed no permeability, showed a permeating flow on some of the days, and only one gave a permeating flow all days during the test period. Only one of the tests, 1H, showed a permeability that was approximately equal to that of the dome salt—about 5.0×10^{-15} cm².

Permeability tests involving the Grand Saline gave consistent values, although they were lower than permeabilities obtained using bedded salt. This difference appeared to be due to the greater relaxation accompanying the disturbance of the purer salt and the greater fragility of the larger crystals in the dome salt.

Since the specimens from the dome formation provided measurable and consistent data, the Klinkenberg constant for helium gas was determined. Relationships between permeability, mean pressure, and liquid pressure were developed, and various factors that reduce permeability were evaluated.

A series of helium tests at different mean pressures were conducted on specimen 11G in determining the Klinkenberg constant. The results of a typical test are shown in Figure 5. A curve of best fit was obtained by using the method of least squares. The computed Klinkenberg constants were 0.19, 0.41, and 0.69, giving an average value of 0.43. Substituting this value into the Klinkenberg formula yielded an equation for the nonreactive liquid permeability, k_L :

$$k_L = \frac{k_G}{1 + \frac{b}{p_m}} \quad (1)$$

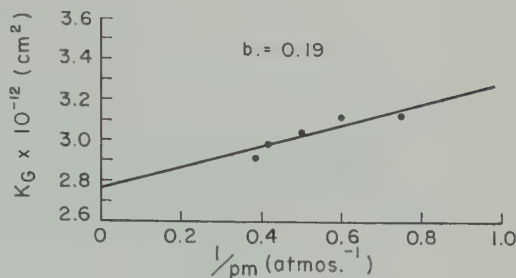


Fig. 5. Klinkenberg data, Grand Saline salt.

k_L = nonreactive liquid permeability.

k_G = permeability as measured with gas.

b = Klinkenberg constant = 0.43.

p_m = mean gas pressure in atmospheres.

The reactivity of kerosene and brine solutions was determined by testing two specimens, 11G and 15G, with helium gas followed by the two solutions. Table 2 shows the results of the tests. The permeability of kerosene, k_K , was 85.5 per cent on the average, based on the nonreactive liquid permeability, k_L ; the permeability of the brine, k_B , averaged 31.5 per cent. This indicates that kerosene, unlike brine, is relatively nonreactive with salt. Kerosene should be nonreactive in the test because it is a nonpolar liquid and is not a solvent for salt. The reactivity of brine with the test specimens is believed to be due to electrokinetic effects which may result from its ionic nature. Another investigator [Washburn, 1922] has shown that saline solutions generate small electromotive forces which cause a retardation of the flow, thus causing the permeability to be smaller than it would otherwise be.

Figure 6 shows the results of all the non-

TABLE 2. Comparison of the Brine, Kerosene, and Nonreactive Liquid Permeabilities of Grand Saline Salt

Confining Pressure P_c , lb/in ²	Average Fluid Pres- sure p_f or p_m , lb/in ²	Mean Stress, σ_m , lb/in ²	Permeability, cm ²					
			k_G	k_L	k_K	k_K/k_L %	k_B	k_B/k_L %
K_G - 1510	10	1500	3.2×10^{-12}	2.3×10^{-12}	2.1×10^{-12}	92	6.2×10^{-13}	27
K_K, K_B - 1600	100							
K_G - 510	10	500	9.0×10^{-11}	7.2×10^{-11}	5.7×10^{-11}	79	2.6×10^{-11}	36
K_K, K_B - 550	50							
Average						85.5		31.5

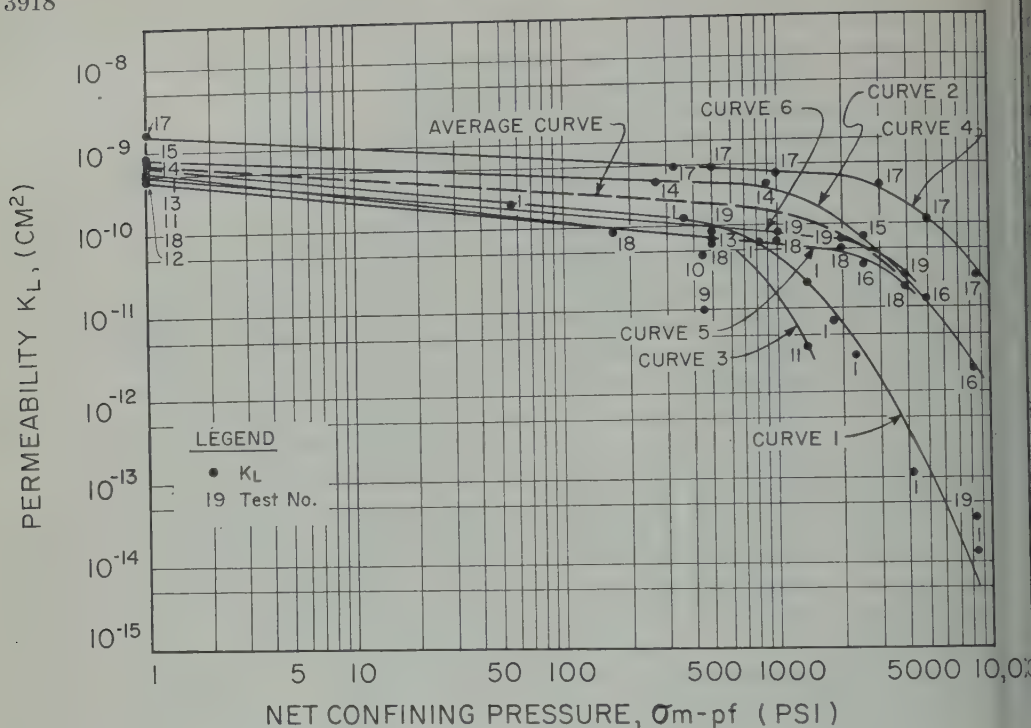


Fig. 6. Nonreactive liquid permeability, Grand Saline salt.

reactive values obtained during the permeability tests. The permeability values at a net confining stress of 1 lb/in² were found to be equal to those obtained from tests at zero overburden pressure.

Curve 1 of Figure 6 represents the average permeability to mineral oil, k_o , for test 1G. Since the permeability to kerosene is approximately equal to the nonreactive liquid permeability, it is assumed that the permeability to the oil will have the same relationship and should be considered equal to the nonreactive liquid permeability. Curve 2 represents the average nonreactive permeability, k_L , for three tests made with helium gas, 14G, 15G, and 16G. The specimens for these tests were made from the same salt block; therefore the results have been incorporated into the same curve. Curve 3 represents the nonreactive permeability for tests 11G and 13G, and these specimens were also made from the same block. Curves 4, 5, and 6 represent the average values for tests 17G, 18G, and 19G. The variation between the results is apparently due to the differences between the individual specimens.

It can be seen from the curves in Figure 6 that the permeability is largely a function of the net confining pressure. When the net confining pressure is between 1 and 1000 lb/in², the following equation is applicable:

$$k_L = C(\sigma_m - p_f)^{-m}$$

$$= 7.1 \times 10^{10} (\sigma_m - p_f)^{-0.212}$$

where

m = slope of the average curve.

C = intercept of the average curve.

k_L = nonreactive liquid permeability (cm²).

σ_m = mean confining stress (lb/in²).

p_f = fluid pressure (lb/in²).

Equation 2 can also be used to describe materials such as sandstones which have a much greater elastic strength than rock salt and do not exhibit considerable plastic flow characteristics at these pressures. The permeability relationship for the net confining pressures above 1000 lb/in² was made by averaging the permeability data collected with these higher pressures. The breakdown of the relationship for larger values of net confining pressure is

to be due to the plastic nature of salt and its low elastic strength.

As noticed that in nearly all of the liquid tests the permeability decreased with the duration of the test and finally reached a minimum value. Table 3 shows the initial and minimum permeabilities obtained in the kerosene and brine tests.

As can be seen from tests 11G and 13G (Table 3) that the nonreactive permeabilities were approximately equal to their respective kerosene permeabilities. However, the nonreactive permeabilities were greater than their brine permeabilities. In both tests the specimens were first tested with helium, then with kerosene, and finally with brine.

In the calcium sulfate and brine test, 3G, and in the synthetic waste test, 4G, were both run at a confining pressure of 950 lb/in². The permeabilities obtained fell within the range of the permeabilities at the same net confining pressure. The permeability values obtained by the brine tests were so varied that no representative equation could be developed; however, an estimation of the brine permeability at a given net confining pressure can be made. The permeability to brine is about 32 per cent of the nonreactive liquid permeability. This is an average taken from Figure 6.

The reduction in permeability appeared to be due to a plugging effect in the upstream end of the specimens. This reasoning is supported by the fact that, when the flow was reversed,

it was observed that there was a subsequent increase in permeability. In test 11G the permeability as measured with kerosene had decreased in 10 days from an initial permeability of 1.9×10^{-12} cm² to 5.4×10^{-13} cm². When the flow was reversed the permeability almost reverted to its initial value. The clogging effect was not anticipated because a membrane filter was used to filter the fluid and thereby remove any debris or solids which might inadvertently have contaminated the fluid.

The effect of a supersaturated brine solution on the permeability was quite drastic. The supersaturated solution was prepared by saturating distilled water at 160°F with sodium chloride and then allowing the solution to cool to the test temperature of 79°F. The supersaturated solution was substituted for the saturated solution on the eighteenth day of the test. The permeability rapidly decreased, and a value of 2.0×10^{-14} cm² was reached on the sixth day after the substitution. The flow was reversed on the sixth day, and the permeability was increased to 6.0×10^{-14} cm². This increase was apparently due to the unclogging of the upstream end of the specimen. It may be noted that the permeability did not return to its original value of 5.0×10^{-13} cm², which existed on the day before the supersaturated solution was used. After the brief rise on the sixth day, the permeability began decreasing again, and it reached zero permeability on the ninth day after the substitution. It should be noted that

TABLE 3. Initial and Minimum Permeabilities of Grand Saline Salt

Confining Pressure P_c , lb/in ²	Average Fluid Pressure P_f , lb/in ²	Mean Stress σ_m , lb/in ²	Permeability	Initial Permeability, cm ²	Min. Permeability, cm ²	Min. $K \times 100\%$ Initial k	Days Duration to Min.
1100	100	1000	k_C	4.4×10^{-12}	1.5×10^{-12}	34.3	8
1100	100	1000	k_S	1.4×10^{-12}	0	0	14
1100	100	1000	k_B	4.5×10^{-12}	1.4×10^{-12}	31.0	3
3200	200	3000	k_B	6.0×10^{-13}	2.0×10^{-13}	3.2	36
2500	200	2300	k_K	5.0×10^{-12}	5.0×10^{-13}	0.001	22
1100	100	1000	k_B	1.0×10^{-12}	2.8×10^{-13}	28.0	4
550	50	500	k_K	1.0×10^{-11}	1.7×10^{-12}	17.0	9
550	50	500	k_K	4.2×10^{-11}	3.2×10^{-11}	76.3	3
1600	100	1500	k_K	2.1×10^{-12}	5.4×10^{-13}	2.1	10
1600	100	1500	k_B	5.0×10^{-13}	0*	0*	9*
550	50	500	k_K	6.2×10^{-11}	1.7×10^{-11}	27.4	6
550	50	500	k_B	4.0×10^{-11}	1.1×10^{-12}	2.7	4

* supersaturated brine solution used.

in this case the blocking of the passages was complete throughout the specimen's length and was not confined to the upstream end, since a reversal of the flow would not re-establish a permeability value near the original value.

After completion of test 16G, in which helium gas was used, the sample was tested at 500 lb/in² net confining pressure using a fine granular suspension in a saturated brine solution. The suspension was prepared by mixing pulverized diatomaceous earth in a saturated solution. It was far from a true suspension, since settling of the suspended particles did occur in 4 or 5 days under quiescent condition. The nonreactive liquid permeability as measured with helium was 2.4×10^{-12} cm². The permeability of the same salt specimen after introduction of the diatomaceous suspension was 4 to 7×10^{-18} cm². This reduction in permeability was about 20 per cent. Furthermore, after 60 minutes, the flow ceased, indicating a complete closure of pores for these conditions.

Porosity and density. Porosity tests were run on eleven Grand Saline and six Hutchinson salt samples, and density tests were run on all but three of the samples. The average porosity and density of the Grand Saline salt specimens were 1.71 per cent and 2.139 g/cm³, respectively, and the maximum deviations from the average porosity and density were 0.54 per cent and 0.20 g/cm³, respectively. Similar values of porosity and density for the Hutchinson salt were 0.59 per cent and 2.153 g/cm³, respectively, and the maximum deviations were 0.17 per cent and 0.01g/cm³. It may be noted that the porosity of the bedded salt was only about one-third that of the dome salt. The slightly greater density of the bedded salt is probably due to its greater impurity content.

There is a definite correlation between permeability and effective porosity for a given medium under different confining pressures. Such a relation is useful in determining one of the two quantities when the other is known. In the analysis of a storage cavity design problem it is necessary to know the porosity of salt at various confining pressures. This could not be determined experimentally because of the complex equipment that would have been required. Therefore, it was computed from permeability values at various confining pressures which were easily measured.

The relation between permeability and porosity is most commonly expressed as:

$$k = C\phi^3/tS_v(1 - \phi)^2$$

where

k = permeability.

C = coefficient depending upon the shape of the cross-sectional area of the pores.

S_v = specific surface area or the ratio of pore's surface area to the mineral volume of the medium.

t = tortuosity factor.

ϕ = effective porosity (%).

The equation is based on Poiseuille's law of flow through capillary tubes and on the assumption that ideal straight capillary tubes exist through the medium. The actual flow paths, however, will not be straight; they will be sinuous and much longer than the ideal pores. The tortuosity factor t compensates for this deviation.

For the same medium under different confining pressures, the factors C , t , and S_v do not vary greatly, and they may therefore be combined into a single constant for this special case.

Combining these factors into a single constant K , gives

$$k = K\phi^3/(1 - \phi)^2$$

which is the equation used herein as the relationship between permeability and porosity.

Since the average permeability and porosity at zero overburden pressure are known, the constant K can be evaluated. Substitution of the permeability value of 7.1×10^{-10} cm² and porosity of 0.0171 gives the constant as 1.37×10^{-4} . Thus the porosity-permeability relationship for the Grand Saline salt is

$$k_L = 1.37 \times 10^{-4} \phi^3 / (1 - \phi)^2$$

Since several assumptions were made in deriving this equation, the relationship should probably be checked for accuracy by porosity-permeability tests before it is actually used in a storage-cavity design.

Conclusions.

1. The storage of radioactive wastes in salt cavities is feasible. Laboratory permeability

The dome salt varied from zero to 1.5×10^{-12} cm², and laboratory permeabilities for the bedded salt varied from zero to 2.3×10^{-13} cm². Although the measured permeabilities are real, the permeability *in situ* must obviously be much less than that measured in the laboratory; the laboratory tests did not involve the extended consolidation that would occur over a geologic time period; also, the laboratory specimens were fractured when they were removed from the mine.

The permeabilities of the bedded salt specimens were smaller than those of the dome salt. In four of the seven tests registered no permeability. Only one specimen having a permeability of 1.2×10^{-13} cm² was comparable to the dome salt.

Bedded salt specimens must be reconsolidated prior to measurement of permeability data. Rock is relatively plastic and elastic, but it relaxes after it is mined. The rapid relaxation and the disturbance due to the mining operations produced localized fractures in the specimens.

Tests made on a solid crystal showed that flow can occur through the crystals themselves.

The effective porosity of the dome salt was 32 per cent; that of the bedded salt was 0.59 per cent.

The nonreactive liquid permeability of the dome salt could be correlated with the net confining stress.

Tests on the dome salt showed kerosene was nonreactive; the brine solutions showed no interaction. The brine solutions had permeabilities which averaged 32 per cent of the nonreactive liquid permeabilities; thus some interaction occurred.

The permeabilities of dome salt to brine (NaCl), calcium sulfate and brine ($\text{CaSO}_4 + \text{NaCl}$), and the synthetic aluminum nitrate waste ($\text{Al(NO}_3)_3 + \text{NaCl}$) were of the order of 2.8×10^{-12} cm². These permeability values were

all approximately equal. All were less than the equivalent nonreactive liquid permeability.

9. Superficial cracks in a cavity can be sealed effectively by the use of a diatomaceous earth suspension. In laboratory tests in which pulverized diatomaceous earth suspended in a brine solution was used, the permeability was reduced to zero within an hour.

Acknowledgments. The Reactor Fuel Waste Disposal Project, of which the research described herein was one phase, was sponsored by the U. S. Atomic Energy Commission. The interest and guidance of Dr. J. Lieberman of the Commission staff is gratefully acknowledged.

REFERENCES

- American Petroleum Institute, *Recommended Practice for Determining Permeability of Porous Media (API RP 27)*, 3rd ed., New York, 1952.
- Heroy, W. B., *Disposal of Radioactive Waste in Salt Cavities*, Report to National Research Council, Division of Earth Sciences, 1956.
- National Academy of Science, *The Disposal of Radioactive Waste on Land*, National Research Council, Report of the Committee on Waste Disposal of the Division of Earth Sciences, Apr., 1957.
- Schechter, J., and E. F. Gloyna, Temperature rise in underground storage sites for radioactive wastes, *Chem. Eng. Progr. Symposium Ser.—Nuclear Eng., Ind. and Eng. Chem.*, 55, 303-310, 1959.
- Serata, S., and E. F. Gloyna, Principles of structural stability of underground cavities, *J. Geophys. Research*, 65, 2979-2987, 1960.
- Serata, S., and E. F. Gloyna, *Development of Design Principle for Disposal of Reactor Fuel Wastes into Underground Salt Cavities*, University of Texas, Sanitary Engineering Research Laboratories Technical Report to the U. S. Atomic Energy Commission, Jan. 1, 1959.
- Washburn, E. W., and E. N. Bunting, Porosity, 5, Recommended procedures for determining porosity by methods of absorption, *J. Am. Ceram. Soc.*, 5(1), 48-56, 1922; Porosity, 6, Determination of porosity by the method of gas expansion, *J. Am. Ceram. Soc.*, 5(2), 112-129, 1922.

(Manuscript received May 13, 1961,
revised August 23, 1961.)

The System $\text{NaAlSi}_3\text{O}_6\text{--H}_2\text{O--Argon}$: Total Pressure and Water Pressure in Metamorphism¹

H. J. GREENWOOD

Geophysical Laboratory

Carnegie Institution of Washington, Washington, D. C.

Abstract. Phase equilibrium in metamorphic rocks is affected by temperature, pressure, the proportions of nonvolatile components, and the chemical potentials of the reacting volatile components. Theory interrelating these variables has been tested by studying the reaction analcite \rightarrow albite + nepheline + water in the presence of mixtures of water and argon. New data on the system $\text{Ar--H}_2\text{O}$ permit calculation of the composition of the water-argon mixture, which should equilibrate with the phase assemblage analcite + albite + nepheline. Experimental determination of this composition as a function of pressure at constant temperature is in good agreement with the theory.

INTRODUCTION

Metamorphic and igneous reactions generally take place in systems that are chemically and mechanically complex. For example, the pressure on the mineral grains will be larger than the partial pressure of any of the components in the fluid phase in contact with the minerals. Pores within a rock may be filled with a fluid at a pressure less than that on the surrounding mineral grains. The rock system may be open or closed with respect to the volatile components, which need not necessarily be present as fluid phases. The understanding of these conditions may be obtained by considering the pressure on the solid phases independently from the partial pressure or chemical potentials of the reacting volatile or mobile components. These two variables, in addition to temperature, have received recent attention in the literature [Danielsson, 1950; Thompson, 1955, 1957, 1959; Yoder, 1952, 1955; Fyfe, Turner, and Verhoogen, 1958; Harker, 1958; Coombs, Ellis, Fyfe, and Taylor, 1959]. With the exception of Harker and Yoder, these writers have used a theoretical approach. This paper reports the results of experiments that agree well with the thermodynamic approach.

THEORY

The theoretical approach used in this paper is similar to that of Thompson [1955], with

Part of a dissertation submitted to the faculty of Princeton University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Geology.

some modification and extension. Theory indicates that mineral equilibria are affected by at least $c_m + 2$ independent variables, where c_m is the number of mobile components. If the components whose relative proportions do not change (fixed or immobile components) are not shown in an equilibrium diagram, then this diagram will need $c_m + 2$ dimensions to represent fully the equilibrium conditions. The observation that two of these dimensions are temperature and pressure leads to the conclusion that the ordinary two-dimensional P - T equilibrium diagram can depict only univariant equilibria having no mobile components. In other words, the system must be closed to all components. If the system should be arranged so that it is open to one of the components, or so that one of the components has its chemical potential controlled by some agent other than the temperature and pressure of the minerals, then another dimension would have to be added to the equilibrium diagram in order to define the chemical potential of that component. This dimension could be taken as chemical potential or any convenient function of chemical potential. The addition of this third dimension changes the ordinary univariant P - T equilibrium line to a three-dimensionally curved divariant equilibrium surface. The detailed shape of this surface for any particular reaction can be calculated thermodynamically and investigated experimentally.

Yoder [1954], using water, and Harker [1958], using carbon dioxide, performed experiments in which the fluid was diluted with argon

and the effect on a phase equilibrium was observed. The expected effects were not observed in Yoder's experiments, probably owing to incomplete mixing of the water and argon. In Harker's experiments the argon and carbon dioxide were premixed, thus avoiding one of the difficulties, but the absence of P - V - T data on the system carbon dioxide-argon makes precise comparison of his results with the predictions impossible.

Thompson [1955, pp. 79-83], in his treatment of open systems, takes as the three main variables P_S , T , and P_F , with the stipulation that P_F be thought of as the pressure on a fluid that is essentially pure, and that can equilibrate with the solid phases. This is similar to the quantity that is here called the equilibrium pressure, P_E . P_E is used instead of P_F to make it possible to restrict the use of P_F to the pressure that acts on any fluid phase, whether pure or not. This notation places no restriction on the system with regard to the purity of the fluid phase or with regard to whether the system is open or closed.

The meaning of $(P_E)_{P_F}^m$ (Table 1) may be better understood by consideration of the following model. The solid mineral phases are completely surrounded by a fluid mixture, with both the fluid mixture and the solid phases under the same hydrostatic pressure. This group of phases is in osmotic equilibrium with one of the pure components of the fluid mixture, component i . Both the mixture of solid and fluid phases and the pure component i are at the same temperature. The pressure of pure component i , separated by a membrane from the solids and the fluid mixture, is equal to $(P_E)_{P_F}^m$. It is now possible to talk without ambiguity of a situation in which the solid phases, at the same pressure as the fluid mixture, are at a higher pressure than the equilibrium pressure of the reacting volatile component, i . Formally, this may be stated

$$P_S = P_F > (P_E)_{P_F}^m \quad (1)$$

It should be noted that the equilibrium pressure is not necessarily equal to the partial pressure.

Any equilibrium reaction involving a single volatile component, i , will be univariant if the pure vapor pressure equals the total pressure ($P_E = P_S$), or if the system is closed to i , but divariant if $P_E < P_S$ and the system is

open to i . This divariant surface will intersect planes parallel to the coordinate planes in curves, the slopes of which are given by the following equations [*Thompson*, 1955, p. 83]:

$$\left(\frac{\partial P_S}{\partial T}\right)_{P_E, i} = \frac{\Delta S}{\Delta V_S}$$

TABLE 1. Terms and Symbols Used in Equations

P_S	The isotropic pressure on the solid phases.
P_F	The pressure on the fluid phase, as distinct from the partial pressure of any of the components in the fluid phase.
$(P_E)_{P_F}^m$	The equilibrium pressure of component i in phase m , where phase m is at the pressure P_F . This is the pressure of pure i which would be in equilibrium with component i in phase m through the medium of a membrane permeable only to component i .
ΔV_S	$(V_S)_{\text{products}} - (V_S)_{\text{reactants}}$.
$(V_i)_{P_E, i}^0$	Volume of pure i given off during a reaction.
S, F	Subscripts; refer to solid and fluid phases respectively [after <i>Thompson</i> , 1955].
μ_i	Chemical potential of component i in phase ϕ .
α	Coefficient of isobaric thermal expansion, where

$$\alpha = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_P$$

β Coefficient of isothermal compressibility, where

$$\beta = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T$$

x_i^m Mole fraction of component i in phase m .

$(f_i)_{P_F}^m$ Fugacity of component i in a phase m which is at a pressure P_F .

$(f_i)_{P_E, i}^0$ Fugacity of pure i at a pressure P_E .

$(\nu_i)_{P_E, i}^0$ Fugacity coefficient of pure i at a pressure of P_E , defined by

$$(\nu_i)_{P_E, i}^0 = \frac{(f_i)_{P_E, i}^0}{P_E}$$

$(\eta_i)_{P_F}^m$ Fugacity ratio of component i in phase m where phase m is at pressure P_F .

$$(\eta_i)_{P_F}^m = \frac{(f_i)_{P_F}^m}{(f_i)_{P_F}^0}$$

η has been called the 'humidity' [*Thompson*, 1955] defined by

$$(\eta_i)_{P_F}^\phi = \exp \left[\frac{(\mu_i)_{P_F}^\phi - (\mu_i)_{P_F}^0}{RT} \right]$$

dP/dT Slope of an equilibrium curve under conditions where $P_E = P_S$.

m Superscript, refers to a one-phase fluid mixture.

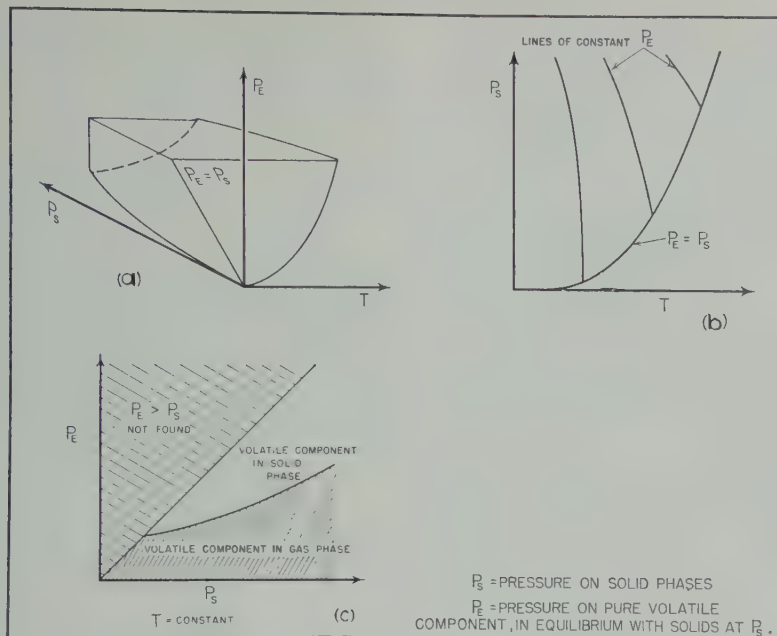


Fig. 1. Diagrams showing the three principal variables: (a) perspective view; (b) projection of (a) onto plane parallel to the P_S - T coordinate plane; and (c) isothermal section through (a).

$$\left(\frac{\partial P_{Ei}}{\partial T}\right)_{P_S} = \frac{\Delta S}{(V_i)_{P_{Ei}}^0} \quad (3)$$

$$\left(\frac{\partial P_{Ei}}{\partial P_S}\right)_T = \frac{-\Delta V_S}{(V_i)_{P_{Ei}}^0} \quad (4)$$

Equations 2, 3, and 4 are modified from Thompson's equations to conform with the notation of Table 1.

The Clapeyron equation for the equilibrium boundary in the presence of the pure² volatile component i may be written, where $P_{Ei} = P_S$,

$$\frac{dP}{dT} = \frac{\Delta S}{\Delta V_S + (V_i)_{P_{Ei}}^0} \quad (5)$$

Equation 2 can be written alternatively as

$$\left(\frac{\partial P}{\partial T}\right)_{P_{Ei}} = \frac{dP}{dT} \times \frac{\Delta V_S + (V_i)_{P_{Ei}}^0}{\Delta V_S} \quad (6)$$

The advantage of writing equation 2 in the form of equation 6 is that if the slope of the equilibrium boundary is known in the presence of the

volatile component can never be strictly pure, as it will always contain some of the coexisting solid phases dissolved in it. These dissolved quantities will be insignificant in the case of most water-rock equilibria except at very high pressures and temperatures.

'pure' volatile component i , and if the volume changes in the reaction are known, a numerical solution may be obtained without reference to other data such as integrated heat capacity equations, or estimated values for the entropy change in the reaction.

Equations 2 to 6 are useful for indicating the general shape of a divariant equilibrium boundary in the dimensions of temperature, pressure on the solid phases, and equilibrium pressure of the reacting volatile component. Figure 1 shows the shape of such a boundary as deduced from these equations. Figure 1a shows the divariant surface in perspective, with the concave side, containing the solid phases with the combined volatile component, closest to the observer. If attention is confined for the moment to conditions under which P_{Ei} equals P_S , only the plane that slopes back in the diagram between the P_{Ei} and P_S need be considered. Conditions in this plane are realized if the solids are in equilibrium with a pure volatile phase at the same pressure. Most hydrothermal experiments closely approach this condition, but never quite achieve it, as the vapor phase is always saturated with the nonvolatile constituents of the run.

Figure 1b shows the same sort of information as Figure 1a but projected onto the P_S - T

co-ordinate plane. Values of constant P_{E_i} are shown as contours, each of which has a negative slope. The negative slope can be accounted for by referring to equation 2 and noting that in general the molar volume of, for example, a hydrous mineral, is greater than that of its anhydrous counterparts, making ΔV_S a negative quantity. In reactions in which ΔV_S is positive, the lines of constant P_{E_i} would have a positive slope. Figure 1b indicates that an increase of P_S independently of P_{E_i} will result in the breakdown of the assemblage containing the volatile component combined in the solid phases (hydrous phases), unless the temperature is lowered at the same time.

Figure 1c shows some of the same information as Figures 1a and 1b, but in a plane of constant temperature, with only one isotherm shown. An increase of P_S at constant P_{E_i} will result in the breakdown of the solid phase containing i , if the temperature is held constant.

Divariant equilibrium boundaries are curved in three dimensions, and before one can make valid calculations based on the approach outlined by equations 2 to 6 this curvature must be determined. In the next section suitable equations are derived.

Curvature of the Univariant Lines

Isotherms. The equation of an isotherm on an equilibrium surface is derived first because much of the experimental work that follows was done at constant temperature. The second derivative of P_{E_i} with respect to P_S may be found by differentiating equation 4 with respect to P_S at constant temperature. The result is expanded by substituting in it

$$\left(\frac{\partial \Delta V_S}{\partial P_S}\right)_T = (\beta V)_r - (\beta V)_p \quad (7)$$

and

$$\left(\frac{\partial(V_i)_{P_{E_i}}^0}{\partial P_S}\right)_T = \left(\frac{\partial(V_i)_{P_{E_i}}^0}{\partial P_{E_i}}\right)_T \times \left(\frac{\partial P_{E_i}}{\partial P_S}\right)_T \quad (8)$$

to give

$$\left(\frac{\partial^2 P_{E_i}}{\partial P_S^2}\right)_T = \frac{(\beta V)_p - (\beta V)_r}{(V_i)_{P_{E_i}}^0} - \frac{(\Delta V_S)^2 \left(\frac{\partial(V_i)_{P_{E_i}}^0}{\partial P_{E_i}}\right)_T}{\{(V_i)_{P_{E_i}}^0\}^3} \quad (9)$$

The compressibilities of the solid phases are much less than that of the evolved fluid than pressures up to about 2000 or 3000 bars equation 9 may be simplified to

$$\left(\frac{\partial^2 P_{E_i}}{\partial P_S^2}\right)_T = - \frac{(\Delta V_S)^2 \left(\frac{(V_i)_{P_{E_i}}^0}{\partial P_{E_i}}\right)_T}{\{(V_i)_{P_{E_i}}^0\}^3}$$

Third-order effects are generally too small to influence results significantly, but they can be evaluated either by an extension of the reasoning leading to equation 10 or by a graphical method. If third-order effects are neglected the equation of an isotherm will be a quadratic in P_S , which might be written as

$$P_{E_i} = \frac{1}{2}C_1 P_S^2 + C_2 P_S + C_3 \quad (10)$$

where

$$\left(\frac{\partial P_{E_i}}{\partial P_S}\right)_T = C_1 P_S + C_2 \quad (11)$$

and

$$\left(\frac{\partial^2 P_{E_i}}{\partial P_S^2}\right)_T = C_1 \quad (12)$$

A numerical solution of equation 11 may be obtained by combining equations 4, 10, 12, and 13 and using values of the volume changes appropriate to the temperature and pressure found from the equilibrium curve where $P_{E_i} = P_S$. This procedure gives a point on the isotherm and its slope and rate of change of slope of the isotherm and thus defines the curve. An example of such a calculation is given in a later part of this paper.

The assumptions of constant rate of change of slope and the small compressibilities of the solid phases relative to the evolved fluid are valid only for modest pressures. At extreme pressures equation 9 should be used in preference to equation 10, and an attempt should be made to evaluate the third derivative. However, for most problems of interest to metamorphic petrology equation 11 will be satisfactory.

Isobars of the equilibrium pressure P_{E_i} . Figure 1b shows the divariant equilibrium surface represented in the form of a projection on the P_S - T coordinate plane. The isobars of P_{E_i} are shown as curved lines. This curvature may be estimated by the same general method used for the isotherms.

Partial differentiation of equation 2 with

et to temperature yields a derivative that be written

$$\left(\frac{\partial S}{\partial T}\right)_{P_{E_i}} = \frac{1}{\Delta V_S} \left[\left(\frac{\partial \Delta S}{\partial T}\right)_{P_{E_i}} - \left(\frac{\partial P_S}{\partial T}\right)_{P_{E_i}} \left(\frac{\partial \Delta V_S}{\partial T}\right)_{P_{E_i}} \right] \quad (14)$$

the expansion of equation 14 it is helpful to note that ΔV_S is not a function of P_{E_i} , making possible to write

$$\left(\frac{\partial \Delta V_S}{\partial T}\right) = \left(\frac{\partial \Delta V_S}{\partial T}\right)_{P_S} + \left(\frac{\partial \Delta V_S}{\partial P_S}\right)_T \left(\frac{\partial P_S}{\partial T}\right)_{P_{E_i}} \quad (15)$$

the definitions of thermal expansion and compressibility allow their substitution in equation 14 giving³

$$\left(\frac{\partial \Delta V_S}{\partial T}\right)_{P_{E_i}} (\alpha V)_p - (\alpha V)_r + \{(\beta V)_r - (\beta V)_p\} \left(\frac{\partial P_S}{\partial T}\right)_{P_{E_i}} \quad (16)$$

the variation of ΔS along the isobar is given by

$$\left(\frac{\partial \Delta S}{\partial T}\right)_{P_{E_i}} = \left(\frac{\partial \Delta S}{\partial T}\right)_{P_{E_i}, P_S} + \left(\frac{\partial \Delta S}{\partial P_S}\right)_{T, P_{E_i}} \left(\frac{\partial P_S}{\partial T}\right)_{P_{E_i}} \quad (17)$$

Equation 17 may be expanded to give

$$\left(\frac{\partial \Delta S}{\partial T}\right)_{P_{E_i}} = \frac{\Delta C_p}{T} + \left(\frac{\partial P_S}{\partial T}\right)_{P_{E_i}} \times [(\alpha V)_r - (\alpha V)_p] \quad (18)$$

Substitution of equations 16 and 18 in equation 14 gives for the rate of change of slope with temperature

$$\left(\frac{\partial S}{\partial T^2}\right)_{P_{E_i}} = \frac{\Delta C_p}{T \Delta V_S} + \frac{1}{\Delta V_S} \left(\frac{\partial P_S}{\partial T}\right)_{P_{E_i}} \cdot \left[2\{(\alpha V)_r - (\alpha V)_p\} + \{(\beta V)_p - (\beta V)_r\} \left(\frac{\partial P_S}{\partial T}\right)_{P_{E_i}} \right] \quad (19)$$

$(\alpha V)_p = (\alpha V)_{\text{products}}$ and $(\alpha V)_r = (\alpha V)_{\text{reactants}}$.

In the event that the heat capacities of all the phases are not known under the conditions of the reaction, a difficulty common to thermodynamic calculations, the ΔC_p may be found from the equilibrium P - T curve where $P_{E_i} = P_S$. The approach is exactly analogous to that used to arrive at equation 19, except that the compressibilities and thermal expansions of the evolved fluid must be included. Inclusion of all the second-order terms gives

$$\frac{\Delta C_p}{T} = \frac{d^2 P}{dT^2} \cdot \Delta V_T - \frac{dP}{dT} \left[2\left\{(\alpha V)_r - (\alpha V)_p - \left(\frac{\partial V_i}{\partial T}\right)_p\right\} + \frac{dP}{dT} \left\{(\beta V)_p - (\beta V)_r - \left(\frac{\partial V_i}{\partial P}\right)_T\right\} \right] \quad (20)$$

which reduces to

$$\frac{\Delta C_p}{T} = \frac{d^2 P}{dT^2} \cdot \Delta V_T + \frac{dP}{dT} \left[2\left(\frac{\partial V_i}{\partial T}\right)_p + \frac{dP}{dT} \left(\frac{\partial V_i}{\partial P}\right)_T \right] \quad (21)$$

if the compressibilities and thermal expansions of the solids are assumed to be much less than that of the evolved fluid. Equation 20 or 21 may be substituted in equation 19 to define the rate of change of slope of an isotherm of the equilibrium pressure, P_{E_i} . If adequate heat capacity data are available for all the phases taking part in the reaction, under the P - T conditions of interest, a cross-check may be obtained between these data and equation 20 or 21.

The equations derived for isotherms and for isobars of P_{E_i} are applied to the experimental data presented in the second part of this paper.

EXPERIMENTAL

Stability of Analcite with $P_{H_2O} = P_S$

This section outlines a redetermination of the equilibrium boundary for the reaction analcite.. \rightarrow albite + nepheline.. + water in the presence of water saturated with analcite, albite, and nepheline. Reversibility of the reaction has been demonstrated at several points, and the curve has been found to lie close to that determined by Sand, Roy, and Osborn [1957], and by Saha [1959].

Most of the experiments were performed in

TABLE 2. X-Ray Data for Synthetic Analcite

<i>I</i>	$2\theta_{\text{CuK}\alpha}$	<i>hkl</i>	d_{hkl}	$Q_{\text{calc}} - Q_{\text{meas}}$ $\times 10^5$
15	48.71	633	1.868	-25
20	47.74	640	1.903	-26
25	40.50	611	2.226	-40
25	37.02	440	2.428	+10
45	35.80	521	2.508	+13
50	33.26	431	2.694	+7
20	31.90	422	2.805	+19
100	30.50	332	2.931	+23
100	25.91	400	3.439	+26
35	24.24	321	3.672	+6
45	18.24	220	4.864	+16
100	15.75	211	5.625	+22

$$d_{(639)} = 78.04^\circ \text{ CuK}\alpha_1$$

$$a_0 = 13.732 \pm 0.002 \text{ \AA}$$

$$\text{Specific gravity (calc)} = 2.113 \text{ g cm}^{-3}$$

standard hydrothermal equipment. Rod bombs of Haynes Stellite 25 were held horizontally in nichrome-wound split furnaces fitted with platinum resistance elements for temperature control. Some of the experiments with pure water were performed in a larger bomb, which was used subsequently with water-argon mixtures. Results obtained with the two types of apparatus are in agreement.

Reagents and starting materials. The silica used for the preparations was cristobalite made by firing quartz from Lisbon, Maryland, at 1350°C for 12 hours and 1500°C for 3 hours. The total content of impurities is 0.03 per cent by weight. The Na₂O was prepared by firing Fisher reagent-grade Na₂CO₃ at 400°C at atmospheric pressure, and continuing to constant weight. It was then weighed in as the carbonate, combined with silica as described by *Schairer and Bowen* [1955], and finally weighed in as the disilicate. The alumina was a certified Fisher reagent, fired to γ alumina.

Starting materials for the experiments were either an oxide mix of analcite composition or the same mix that had been crystallized under hydrothermal conditions to either analcite or albite + nepheline. The oxide mix was prepared by grinding together for 2 hours in a motor-driven agate mortar the proper proportions of sodium disilicate, alumina, and cristobalite. The preparation of albite + nepheline required only a few days at temperatures above the breakdown temperature of analcite under hydrother-

mal conditions. The preparation of analcite required runs of about 3 weeks in duration at temperatures well within the field of stability of analcite. Even then it was possible to detect the presence of small amounts of albite on X-ray diffraction patterns. The appearance and persistence of albite in the analcite stability field is attributed to a more rapid rate of nucleation of albite than of analcite, and to the slower rate of conversion of crystalline albite to analcite by reaction with nepheline and water. The presence of minor amounts of albite in rocks otherwise consisting of analcite indicates that the analcite was slightly deficient in silica, compared with natural specimens, which often have more than the stoichiometric amount of silica.

Phases synthesized and their stability relations. X-ray data for analcite are presented in Table 2. The reflection angle for the (639) plane of synthetic analcite is at $2\theta = 78.04^\circ$ (CuK α_1), determined by comparison with the (331) reflection of silicon, which was used as an internal standard. *Saha* [1959, Fig. 2] presents a determinative curve for the amount of silica in analcite, based on this reflection. This curve indicates that the analcite synthesized in the present work has a molecular ratio of SiO₂:Na₂O between 3.5 and 3.6. This may be compared with the ideal stoichiometric ratio of 4. Analcite grown in open tubes, in sealed tubes, and in the presence of argon and water all have the same spacing of the (639) planes, and it is presumed on this basis that they all have the same silica content.

Saha [1959] also reported that the water content of analcite varies with its silica content. His Figures 2 and 4 indicate that the molecular ratio of H₂O to Na₂O is 1.87 in the analcite synthesized here. This was checked by performing weight-loss analyses for water on the samples of synthetic analcite, one of which had been grown in the presence of an argon-water mixture. All three had the same proportion of water, 8.6 ± 0.2 per cent by weight, corresponding to a water/soda ratio of 2.0 ± 0.05 . The discrepancy between the ratio according to the curve of *Saha* and the ratio as determined in this study corresponds to about 0.4 mg in an average sample weight of 70 mg, and may easily not be a real difference. Accordingly, the analcite is assumed to be of the composition indicated by the data of *Saha*, and the consequences

ing otherwise are evaluated at a later in this paper. The composition of the albite is therefore considered to be $\text{Na}_{1.07}\text{Si}_{1.93}\text{O}_8\text{H}_2\text{O}$.

te added in proof. Saha [1961] has re-published a correction to his 1959 determinative curve for the composition of analcite. A new curve indicates that the analcite synthesized here is slightly more sodic than was assumed on the basis of the 1959 curve. The change is within the uncertainty of the determinative curve and well within the uncertainty of the calculations made in this paper.

albite and nepheline formed 100 per cent of runs of all starting materials held for more than 3 days at more than 30°C above the albite breakdown temperature. The albite is identical with natural and synthetic high albite, and is compared with heated Amelia albite [Smith, 1956] in Table 3.

The nepheline synthesized in this study has a composition intermediate between NaAlSiO_4 and NaSi_3O_8 , owing to omission solid solution of alkali in the nepheline [Donnay, Schairer, and Taylor, 1959]. The composition of the nepheline was not measured in this study. The unit cell of the mineral does not change appreciably with changes in the alkali content along the nepheline-albite join, and the composition was therefore taken from the work of Hamilton [1961, Fig. 2]. This indicates that the nepheline solid solution

TABLE 4. X-Ray Data for Synthetic Nepheline

I	$2\theta_{\text{CuK}\alpha}$	hkl	d_{hkl}	$Q_{\text{calc}} - Q_{\text{meas}} \times 10^5$
15	47.00	32.1	1.926	+111
15	45.67	32.0	1.986	+ 95
20	43.24	31.2	2.092	+269
25	39.078	31.1	2.302	+ 7
10	38.41	20.3	2.343	- 16
15	37.51	30.2	2.398	+359
20	34.838	21.2	2.575	- 2
50	31.023	30.0	2.883	+ 13
100	29.713	20.2	3.007	+ 1
100	27.300	21.0	3.267	0
100	23.173	20.1	3.838	- 9
75	21.270	00.2	4.177	- 25
50	20.57	11.1	4.317	- 77
25	17.80	11.0	4.981	- 14

$$a_0 = 9.982 \pm 0.005 \text{ \AA}$$

$$c_0 = 8.373 \pm 0.005 \text{ \AA}$$

$$\text{Specific gravity} = 2.612 \text{ g cm}^{-3} \text{ (calc)}$$

contains 15 weight per cent of the albite molecule at 500°C. The nepheline formula can therefore be written $\text{Na}_{7.85}\square_{0.64}\text{Al}_{7.35}\text{Si}_{3.64}\text{O}_{33}$ where \square indicates vacant alkali sites. An indexed powder pattern appears in Table 4.

Paragonite was observed in one 3-day run that was cooled slowly to room temperature from 500°C at a pressure of 500 bars. It was present in amounts too small to appear on the X-ray diffractometer chart, and was identified tentatively on the basis of its high birefringence and micaceous appearance. Another solid phase was observed in some runs that were quenched by suddenly releasing the pressure and simultaneously plunging the bomb into cold water. It was found only in runs made with sealed capsules and seemed not to depend critically on the temperature. There was not enough of the material to appear on X-ray diffraction charts, and its optical properties were not sufficient to identify it. It is neither tridymite nor cristobalite, nor is it apparently any of the likely zeolitic minerals, except perhaps zeolite 'B' of Saha, which has the same refractive index. It has minimum and maximum refractive indices of 1.482 and 1.495, respectively, is length-fast, and has an extinction angle of $X' \wedge$ elongation of 4°. It is thought to play no part in the phase equilibria studied.

The chemical reaction representing the stability of analcite solid solution in equilibrium

TABLE 3. X-Ray Data for Synthetic Albite

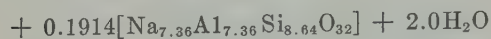
$2\theta_{\text{CuK}\alpha}$	hkl	d_{hkl} Measured	d_{hkl} Heated Amelia
5	42.67	003	2.119
5	42.17	060	2.143
5	35.96	312	2.497
0	35.54	241	2.526
0	33.75	132	2.657
5	31.56	131	2.835
5	30.296	041	2.9497
0	27.785	040	3.2097
5	24.44	130	3.642
0	23.76	130	3.745
5	22.03	201	4.035

$$2\theta_{131}(\text{CuK}\alpha) = 2.19^\circ \text{ (synthetic).}$$

$$2\theta_{131}(\text{CuK}\alpha) = 2.16^\circ \text{ (heated Amelia).}$$

$$\text{Specific gravity} = 2.606 \text{ g cm}^{-3} \text{ (calculated from unit cell of Smith, 1956).}$$

with albite, nepheline solid solution, and water can be written



The equilibrium data for this reaction in the presence of water saturated with analcite, nepheline, and albite are presented in Table 5, and are shown graphically in Figure 2. The heavy circles indicate the runs in which the reversibility of the reaction has been demonstrated, and the lighter circles indicate the runs in which oxide mixes on the stoichiometric analcite composition crystallized to the indicated phases.

P-V Data on the System Water-Argon at 500°C

The interpretation of phase-equilibrium obtained in the presence of mixtures of water and argon requires data on the system water-argon. This system has been investigated at 500°C in the pressure range between 500 and 2000 bars, and the fugacity of water in mixtures has been calculated from the results of *P-V* data.

Apparatus and measurement. The bomb used for the experiments is illustrated in Figure 1. It was made from a chromium-nickel-cobalt alloy (Allegheny-Ludlum S-816). The bomb has a capacity of $9.65 \pm 0.08 \text{ cm}^3$ at 25°C and 1 atmosphere, as determined by weighing

TABLE 5. Results of Experiments in the System $\text{NaAlSi}_2\text{O}_6\text{-H}_2\text{O}$

Pressure, bars	Temperature, °C	Starting Materials	Duration, hours	Products
1000	545	ox	27	Ab + Ne
1010	544	ox	34	Ab + Ne
1000	540	ox	19	Ab + Ne
1000	522	ox	21	Anc + (Ab + Ne)
1980	566	ox	20	Anc + (Ab + Ne)
2000	573	ox	37	Ab + Ne
495	511	ox	66	Ab + Ne + (Anc)
495	511	ox	66	Ab + Ne + (Anc)
1000	550	ox	24	Ab + Ne
1050	493	Ab + Ne	115	Anc + (Ab + Ne)
1500	545	ox	81	Ab + Ne + Anc
1500	559	ox	81	Ab + Ne
1500	500	Ab + Ne + Anc	140	Anc + (Ab + Ne)
1500	536	ox	140	Anc + (Ab + Ne)
500	520	ox	161	Ab + Ne
500	501	ox	137	Ab + Ne
1000	536	ox	117	Ab + Ne
1000	526	ox	117	Ab + Ne
1000	516	ox	153	Anc + (Ab + Ne)
255	451	ox	158	Anc + (Ab + Ne)
255	477	ox	158	Ab + Ne
500	493	Anc + (Ab + Ne)	332	Ab + Ne + (Anc)
500	482	Anc + Ab + Ne	332	Anc + (Ab + Ne)
125	431	Ab + Ne	421	Ab + Ne
125	458	Anc + (Ab + Ne)	425	Ab + Ne
135	430	Anc + (Ab + Ne)	429	Ab + Ne + (Anc)
125	400	Ab + Ne	312	Anc + (Ab + Ne)
700	502	Ab + Ne	504	Ab + Ne + (Anc)
595	498	Anc + (Ab + Ne)	504	Ab + Ne
2000	558	Ab + Ne	528	Anc + Ab + Ne
2000	577	Anc	432	Ab + Ne + (Anc)

ox, oxide mix.

Ab + Ne, albite plus nepheline, bulk composition $\text{NaAlSi}_2\text{O}_6$, crystallized hydrothermally.

Anc, analcite, crystallized hydrothermally.

(Ab + Ne), minor proportions of albite plus nepheline.

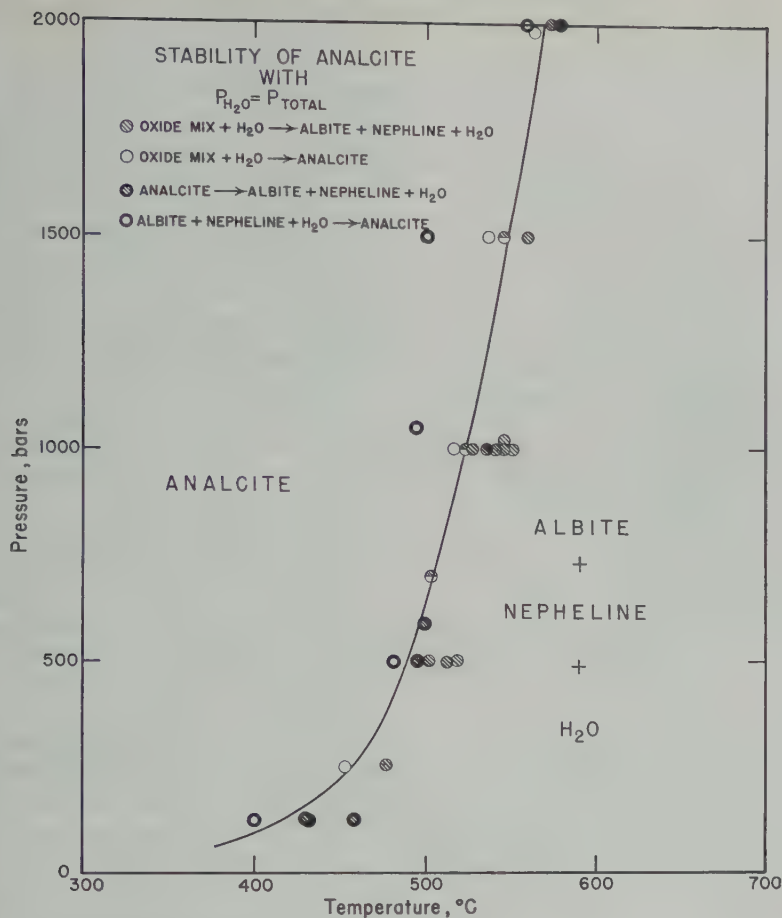


Fig. 2. Stability of analcite in the presence of water saturated with $\text{NaAlSi}_3\text{O}_8$.

ant of contained mercury, filling the bomb with argon at room temperature and pressures to 2000 bars, and calculating the volume of the bomb from the data of Michels, Wijker, and Wijker [9]. Agreement between the methods was within the stated uncertainty.

The bomb is connected to the pressure line by a capillary stem 18 inches long, made of stainless steel pressure tubing with a measured volume of 0.155 cm^3 . The stem and bomb are isolated from the rest of the system by a pressure valve having a calculated volume of 0.17 cm^3 . When the contents of the bomb are expelled into the analysis apparatus, the contents of the stem and valve are also expelled. The amount of material from the stem and valve is computed from their volumes, the pressure in the line, and the temperature distribution between the bomb and the valve, and this

quantity is subtracted from the total found by analysis.

Corrections were applied to the measured volume of the bomb at high temperature and pressure to account for thermal expansion and elastic stretch, which contributed 2.0 per cent and 0.34 per cent, respectively.

Pressure was measured with a calibrated Ashcroft bourdon tube gage, graduated from 0 to 40,000 psi. The pressure in the bomb could be read with a maximum uncertainty of about 10 bars.

Temperature was measured with chromel-alumel thermocouples, which were calibrated against the melting points of cesium chloride, sodium chloride, and zinc. The temperature calibration of the bomb was done by comparing the reading of the thermocouple in the well with the reading of a calibrated thermocouple probe in-

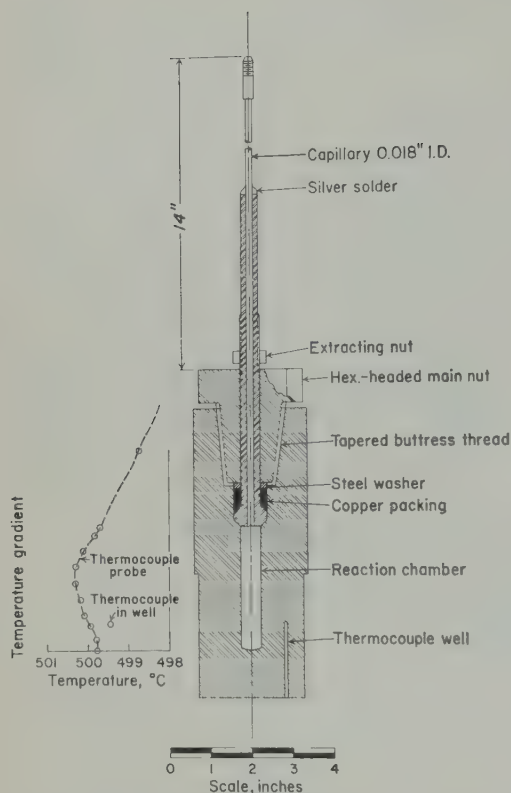


Fig. 3. Bomb used for P - V - T measurements in the system water-argon, with graph of thermal gradient.

serted through a special head and stem. The results appear in Figure 3. The gradient is small, with the temperature in the bomb at all points within half a degree of the average. The emf of the thermocouple was measured with a Leeds and Northrup White potentiometer. Temperature readings are reproducible to better than 0.5°C and are probably accurate to within about 2.0°C .

The apparatus used for the analysis of the water-argon mixtures is shown in Figure 4. It consists of two parts, the condenser for the water vapor (cold trap) and a 10-liter bottle to contain the argon. All the apparatus is of Pyrex glass with conventional stop-cocks and ground-glass joints. The volume of the whole analysis system was measured as $9712 \pm 6 \text{ cm}^3$, by expanding weighed amounts of argon into it at low pressure and calculating the volume from the tabulated data of *Hilsenrath and others* [1955].

The bomb is filled while cold with cold distilled water in order to exclude air and is connected to a reservoir containing water at a pressure of about 500 bars. The bomb is placed in a vertical nichrome-wound furnace and allowed to heat up to the temperature of the run. Argon at high pressure is then allowed to enter the bomb. Pressure gradients in the lines vanish after about 10 seconds.

The contents of the bomb are analyzed for water and argon by removing the bomb from the furnace and releasing the gas into an evacuated analysis apparatus. The cold trap is immersed in a sludge of solid carbon dioxide and acetone, and the amount of water recovered is determined by weight difference. Adiabatic cooling occurs at the junction of the capillary and the cold trap, keeping the glassware cool. The argon passes through the porous plug of glass wool in the cold trap and is collected in the 10-liter bottle, where its pressure is measured with the mercury manometer.

Homogeneity of the water-argon mixture is established by releasing a portion of the gas into the analysis system and making an analysis, and then releasing the rest into the system. There was no significant difference between the proportions of argon and water in the two fractions of the same run.

TABLE 6. Isotherm of Water, 500°C

Pressure, bars	Volume, $\text{cm}^3 \text{ mole}^{-1}$
53	753.0
154	350.0
179	476.0
200	250.0
259	204.0
300	148.2
325	157.8
385	109.0
400	99.3
420	92.4
490	75.2
500	69.4
500	71.8
500	71.0
500	67.9
529	66.2
631	55.1
650	48.4
1000	34.6
1500	28.4
1500	28.4

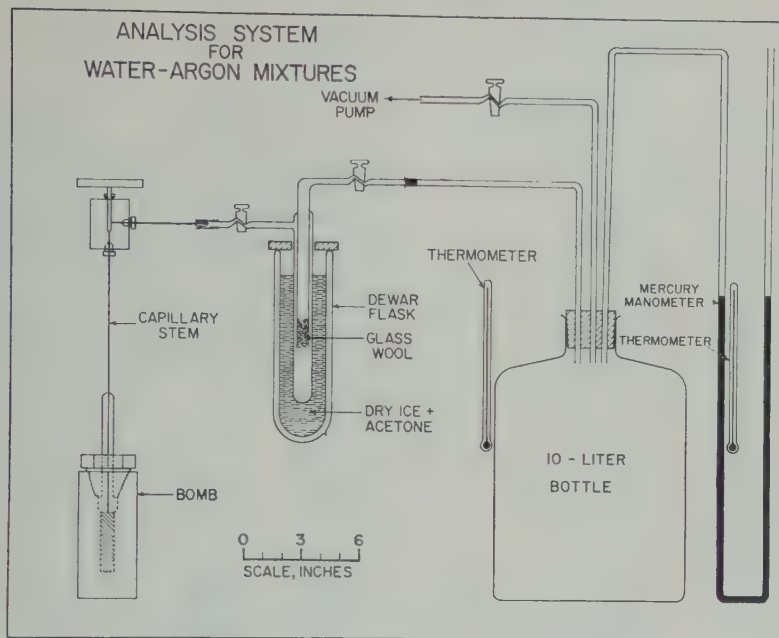


Fig. 4. Analysis apparatus for system water-argon.

Results. The P - V - T data secured by the above method are in excellent agreement with data obtained by other workers by quite different methods. The data for pure water and for pure argon are given in Tables 6 and 7, respectively,

and are shown graphically in Figures 5 and 6. The size of the circles plotted in the figures indicates the estimated uncertainty in the quantities, and the curves drawn with light lines represent the data of Kennedy [1950], Holser and

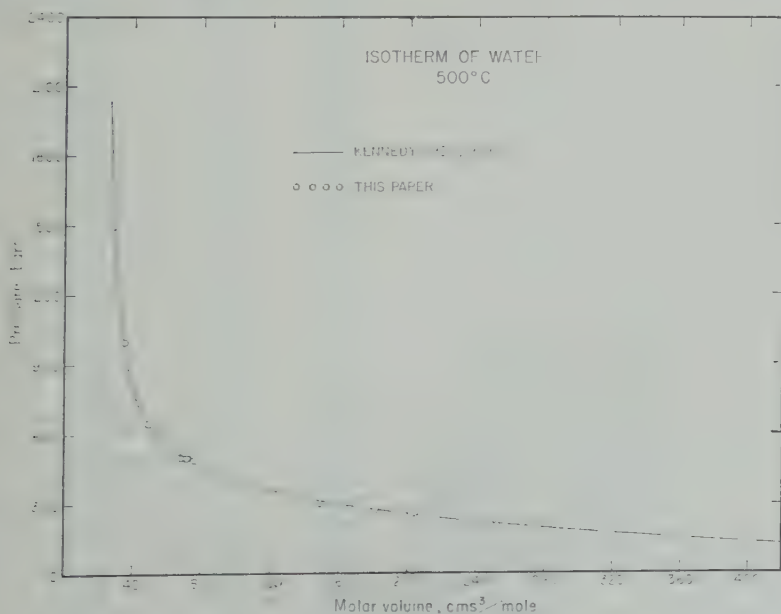


Fig. 5. 500°C isotherm of water.

TABLE 7. Isotherms of Argon

Pressure, bars	Volume, cm ³ mole ⁻¹
150°C	
239	178
380	105.6
487	86.2
617	71.2
781	60.9
1097	50.4
1422	43.6
1841	39.5
500°C	
305	220.7
560	132.8
820	95.6
1085	78.1
1350	66.8
1485	62.5
1620	59.3
1755	56.3
1890	54.2
2025	52.1

Kennedy [1959] (Fig. 5); and Michels, Wijker, and Wijker [1949] (Fig. 6). Two isotherms of argon are shown in Figure 6, the one at 150°C being the highest temperature data available with which to check this work. The heavier line, for the 500°C isotherm, is drawn from the new work presented here (cf. Table 7).

The data for mixtures of water and argon are given in Table 8, in terms of pressure, molar volume, and composition, at 500°C. The data in this table have not been smoothed or altered in any way from the values actually measured, except to account for the thermal expansion and pressure dilation of the pressure vessel. The data are plotted in Figure 7, in dimensions of total pressure and mole fraction, and contoured in terms of total molar density.

The data are given a conventional thermodynamic treatment, assuming that no compounds or gas species intermediate between argon and water are formed. The fugacity of a component of a mixture is obtained by integrating an excess volume function found from the data on the mixture. The method is described in summary by Bennett [1953]. The quantity obtained most directly from isothermal data on the mixture is a fugacity ratio, which is given by the following equation:

$$\ln \frac{(f_i)_{P_F}^m}{x_i(f_i)_{P_F}^0} = \frac{1}{RT}$$

$$\cdot \int_0^{P_F} [(\bar{V}_i)_{P_F}^m - (\bar{V}_i)_{P_F}^0] dP_F \quad (1)$$

Equation 22 is valid for constant temperature and mole fraction x_i . The integrand in equation 22 is more than zero in the limit of zero pressure [Gillespie, 1929]. The limiting value is given by

TABLE 8. Isotherm of System
Water-Argon, 500°C

Pressure, bars	Mole fraction of water, x_{H_2O}	Volume cm ³ mol ⁻¹ , \bar{V}^m	Volume change on mixing $\Delta \bar{V}^m$
500	0.446	144.3	30.92
500	0.639	132.6	33.92
500	0.759	116.0	26.6
500	0.875	101.9	21.4
514	0.749	114.3	27.23
800	0.322	96.3	17.18
800	0.530	87.3	20.44
800	0.652	78.7	18.96
800	0.881	56.0	9.76
915	0.796	57.1	9.65
1100	0.557	68.7	16.4
1110	0.203	78.0	10.30
1110	0.674	57.6	10.70
1110	0.872	43.1	5.19
1112	0.887	39.3	1.66
1200	0.456	65.9	10.74
1200	0.571	58.4	8.25
1200	0.578	58.6	8.70
1215	0.862	40.3	3.27
1350	0.345	62.6	6.88
1380	0.854	37.4	2.30
1494	0.460	55.8	7.65
1500	0.328	60.5	8.24
1506	0.503	53.2	6.63
1506	0.631	49.0	7.20
1655	0.554	48.2	5.98
1800	0.099	56.6	1.86
1930	0.205	51.9	3.44
1960	0.465	47.2	5.43
1965	0.526	42.6	2.77
1965	0.528	42.8	2.98
1993	0.311	50.9	5.23
1993	0.354	49.5	5.00
2000	0.200	51.0	2.21
2000	0.212	51.8	3.29
2000	0.393	48.7	5.33
2000	0.419	47.5	4.90
2000	0.604	39.3	2.02
2000	0.892	28.9	-0.15
2000	0.950	26.4	-0.99
2010	0.601	39.5	2.20

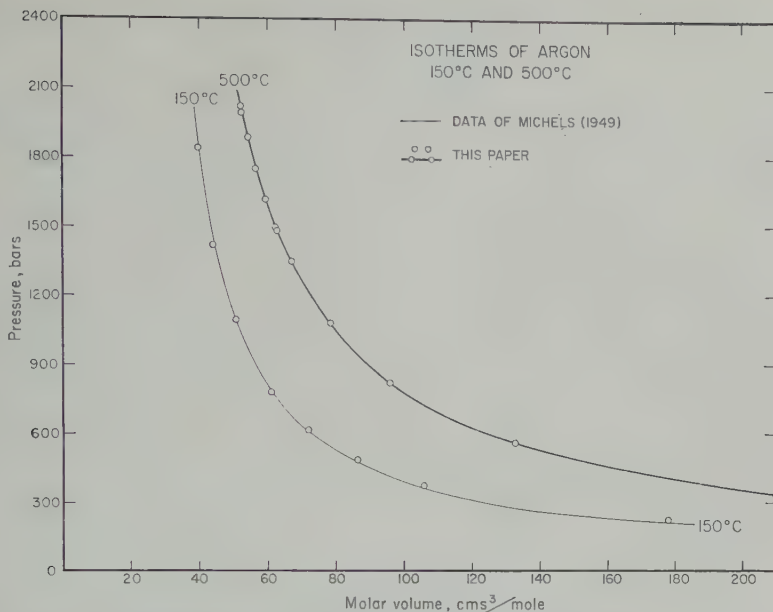


Fig. 6. 150°C and 500°C isotherms of argon. Data of *Michels, Wijker, and Wijker* [1949] shown by light line; new data, by circles and heavy line.

$$[(\bar{V}_1)_{PF}^m - (\bar{V}_1)_{PF}^0]$$

$$= \frac{1}{RT} (x_2)^{1/2} [(A_1)^{1/2} - (A_2)^{1/2}] \quad (23)$$

a binary mixture of species 1 and 2, having
al coefficients for the cohesive pressure of

A_1 and A_2 for pure 1 and pure 2, respectively. If a precise equation of state from which to obtain A_1 and A_2 is not available, only slight error is introduced by using values from the van der Waals equation, derived from critical data.

Equation 22 cannot be evaluated until the

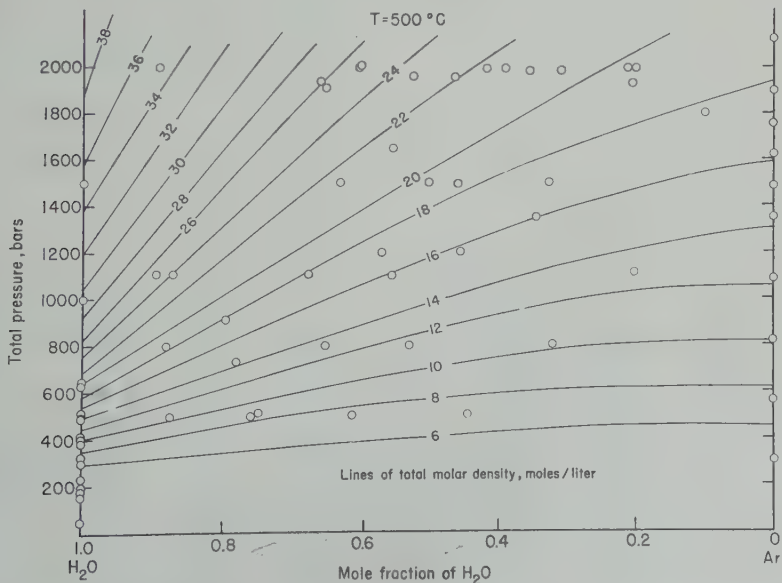


Fig. 7. Isotherm of the system water-argon, showing isopleths of constant molar density.

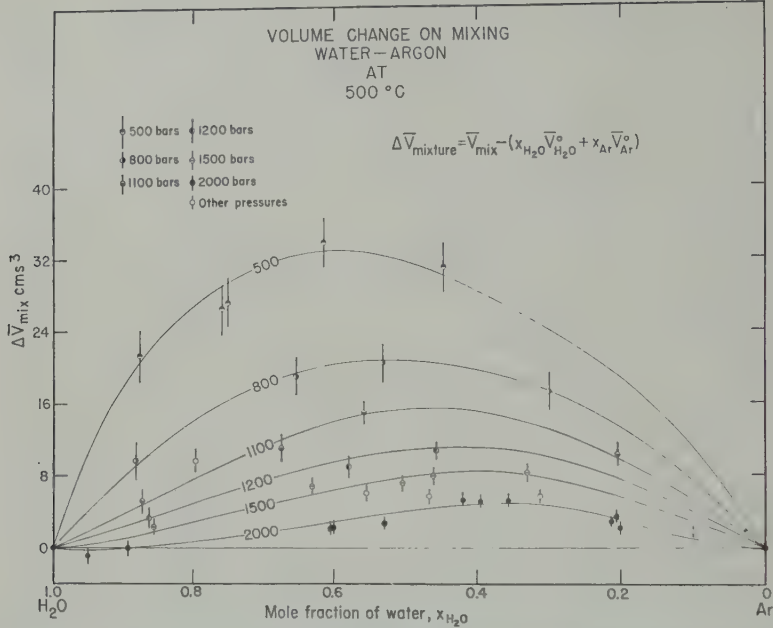


Fig. 8. Volume change on mixing in the system water-argon at 500°C, showing isopleths of constant pressure.

value of the integrand as a function of pressure at constant composition is known. The integrand is given by

$$[(\bar{V}_1)_{P_F}^m - (\bar{V}_1)_{P_F}^0]$$
$$= (\Delta \bar{V}^m)_{P_F} - x_2 \left(\frac{\partial \Delta \bar{V}^m}{\partial x_2} \right)_{P_F, T} \tag{24}$$

where

$$(\Delta \bar{V}^m)_{P_F} = (\bar{V}^m)_{P_F} - [x_1(\bar{V}_1)_{P_F}^0 + x_2(\bar{V}_2)_{P_F}^0]_{P_F, T} \tag{25}$$

The volume change on mixing, $(\Delta \bar{V}^m)$, calculated from equation 25, is given in Figure 8 for pressures ranging from 500 to 2000 bars. (See also the last column of Table 8.) The results are contoured in terms of constant pressure on the gas, P_F . Equation 24 indicates that

TABLE 9. Fugacity Ratio of Water in the System Water-Argon at 500°C

$$\eta_{H_2O} = \frac{f_{H_2O}^m}{f_{H_2O}^0}$$

Pressure, bars	Mole fraction of water, x_{H_2O}								
	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
200	0.895	0.785	0.669	0.581	0.494	0.407	0.320	0.215	0.122
400	0.919	0.842	0.747	0.667	0.582	0.481	0.379	0.270	0.154
600	0.927	0.871	0.809	0.742	0.666	0.567	0.455	0.331	0.194
800	0.941	0.892	0.846	0.789	0.728	0.637	0.529	0.397	0.240
1000	0.947	0.903	0.863	0.819	0.771	0.692	0.588	0.460	0.288
1200	0.948	0.907	0.865	0.827	0.791	0.724	0.634	0.517	0.338
1400	0.951	0.909	0.868	0.835	0.803	0.741	0.669	0.564	0.382
1600	0.949	0.908	0.868	0.835	0.804	0.754	0.693	0.602	0.581
1800	0.947	0.904	0.861	0.836	0.806	0.762	0.714	0.594	0.412
2000	0.948	0.906	0.871	0.838	0.807	0.772	0.730	0.667	0.480

Standard state is water at the pressure and temperature of the mixture.

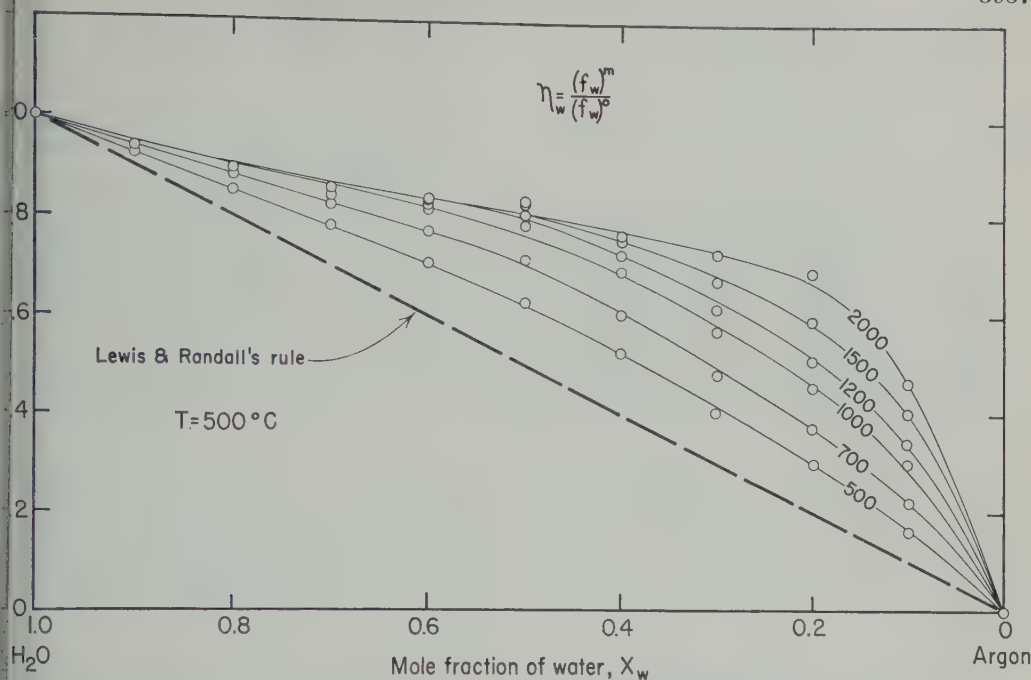


Fig. 9. Comparison of water-argon mixtures with Lewis and Randall's assumption of ideal mixing.

method of obtaining the volume change of mixing for water from Figure 8. Figure 8 is a very sensitive test of the internal consistency of the data than Figure 7, because $\Delta \bar{V}^m$ is a small difference between large numbers.

The integration in equation 22 is performed at constant composition on curves of $(\Delta \bar{V}_{H_2O})^m$ versus pressure, which were derived from Figure 8 by application of equations 23 and 24. The curves were integrated with a planimeter, reduced to the fugacity ratio $\eta_{H_2O}^m = (f_{H_2O})^m / (f_{H_2O})^0$, and plotted in Figure 9. Values of $\eta_{H_2O}^m$ are taken from a large-scale plot of Figure 9 and are given in Table 9. Figure 9 facilitates the comparison of the argon-water data with the rule of Lewis and Randall, which states that the fugacity ratio should be equal to the mole fraction. The rule is satisfactory only at pressures less than 500 bars.

The compressibility factor, $Z = PV/RT$ (Table 10), was calculated after graphically smoothing the data on large-scale plots of molar density versus composition and molar density versus pressure. Departures of the calculated values from the measured values are within the experimental error.

Accuracy of the calculated quantities. The

measurements which lead to the number of moles of argon and of water in the bomb are accurate to about 1 per cent, so that the uncertainty in the mole fraction is estimated to be about 2 per cent. Molar volumes of the mixtures are calculated from the measured contents and volume of the bomb. The estimated uncertainty in each of these figures is 1 per cent, and it is estimated that the uncertainty in the resulting molar volume is about 2 per cent. The estimate is supported by the agreement of the data with those of *Michels, Wijker, and Wijker* [1949], and of *Kennedy* [1950] on pure argon and pure water.

The volume change on mixing (cf. Table 8 and Fig. 8) depends on all the measured quantities in a critical way, because it ranges between 0 and 25 per cent of the measured molar volumes. A 2 per cent error in the molar volume of the mixture could appear as a 10 or 15 per cent error in the volume change term. This high possible error seems excessive because the internal consistency is generally within 3 per cent. The uncertainty in the volume change on mixing is thus estimated to be between 3 and 5 per cent.

The percentage uncertainty in the fugacity ratio is necessarily the same as that of the volume change on mixing and is also taken to be about

TABLE 10. Compressibility Factor for
Water-Argon Mixtures at 500°C
 $Z = PV/RT$

Pressure, bars	Mole Fraction of Water in Mixture					
	1.0	0.8	0.6	0.4	0.2	0.0
400	0.640	0.864	1.00	1.07	1.13	1.15
600	0.501	0.812	1.00	1.10	1.19	1.20
800	0.492	0.790	1.01	1.15	1.13	1.27
1000	0.531	0.802	1.03	1.21	1.29	1.34
1200	0.582	0.830	1.06	1.24	1.34	1.41
1400	0.633	0.861	1.10	1.29	1.41	1.48
1600	0.688	0.899	1.14	1.33	1.46	1.55
1800	0.746	0.943	1.19	1.39	1.53	1.62
2000	0.808	0.994	1.24	1.45	1.59	1.70

5 per cent. It must be appreciated that the absolute possible error in these last-mentioned quantities could be as high as 15 per cent, although the internal consistency suggests a much lower error.

Stability of Analcite with $P_{\text{H}_2\text{O}} < P_s$

The addition of argon to the system $\text{H}_2\text{O}-\text{NaAlSi}_3\text{O}_8$ displaces the equilibrium boundary between the stability fields of analcite and albite + nepheline. The displacement is due to the combined effect of diluting the water and elevating the total pressure through the addition of argon. The equilibrium pressure of water defined by the equilibrium boundary between analcite and albite + nepheline may be calculated from the above equations. It may be compared with the equilibrium pressure of water in the water-argon mixtures in equilibrium with the assemblage.

The equilibrium pressure of water in the water-argon mixtures was obtained from a large-scale plot of Figure 10. This figure was plotted from the data in Table 9 and the fugacities given by *Holser* [1954]. It provides an easy method of visualizing and calculating the equilibrium pressure of water in a water-argon mixture. The equilibrium pressure of water in the mixture is obtained by constructing a horizontal line from the appropriate mole-fraction line at the appropriate total pressure to the line for pure water, and reading off the pressure of the pure water. For example, in a mixture containing 40 mole per cent water at a pressure of 1800 bars the equilibrium pressure of water is 1200 bars. The uncertainty in the equilibrium pressure thus derived is assumed to

be about 5 per cent, in accordance with the uncertainty estimated for the calculated fugacities.

Equilibrium pressure is not the only function of the chemical potential of water that might be used to define equilibrium conditions. An alternative, suggested by *Thompson* [1957], is the fugacity ratio, which may in the case of water be termed the 'humidity.' The humidity $\eta_{\text{H}_2\text{O}}$ related to the fugacity and the equilibrium pressure by the following identity:

$$(\eta_{\text{H}_2\text{O}})_{P_F}^m \equiv \frac{(f_{\text{H}_2\text{O}})_{P_F}^m}{(f_{\text{H}_2\text{O}})_{P_F}^0} = \frac{(\nu_{\text{H}_2\text{O}})_{P_F, \text{H}_2\text{O}}^0 \cdot P_{\text{H}_2\text{O}}}{(\nu_{\text{H}_2\text{O}})_{P_F}^0 \cdot P_F} \quad (2)$$

(P, T const)

In discussions of metamorphic systems the choice between using P_s , f , μ , or η is dictated by convenience, as only one of the four can be independent at constant temperature, and all are equally applicable to both open and closed systems.

Apparatus and measurement. The apparatus used was the same as that used for the system water-argon, except for the containers for the solid phases. The samples of solid phases were held in small open platinum envelopes supported on a gold rack. Four charges of about 15 mm each were used in each run, two at the top of the bomb and two at the bottom. The pairs of samples at the top and bottom were identical, each pair consisting of one sample of albite + nepheline and one of analcite with minor albite and nepheline. The purpose of using four charges was to demonstrate reversibility in every run, and to test the homogeneity of the gas phase by comparing the products of the reaction from the capsules at the top with those at the bottom. All capsules gave consistent results.

Phases synthesized. The physical properties of the phases synthesized in the presence of argon are identical with those of the phases synthesized in the presence of nearly pure water. The X-ray data indicate that no large structural changes in the phases resulted through the use of argon, and it seems that no appreciable free-energy changes resulted, or that equilibrium would probably have been displaced in a way not predicted by theory.

A synthetic analcite grown in the presence of argon has been analyzed for argon by L.

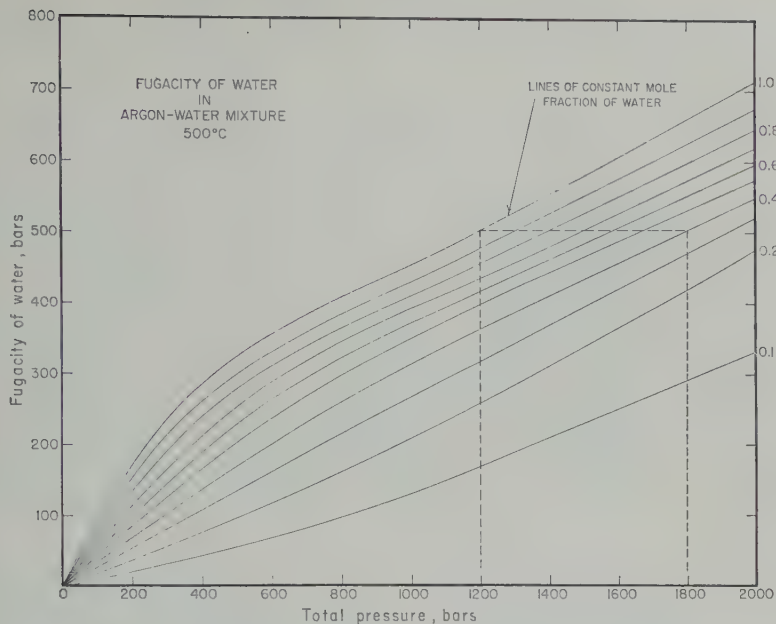


Fig. 10. Fugacity of H_2O in the system $\text{H}_2\text{O--Ar}$ at 500°C .

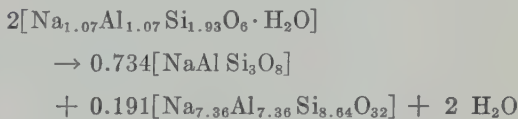
rich of the Department of Terrestrial Magnetism, Carnegie Institution of Washington, by isotope dilution procedure used in analyzing on from rock-forming minerals. The analcite contained 0.025 ± 0.005 per cent by weight of on.

The fluid phase in equilibrium with the solid phases has been analyzed for its sodium content by flame photometry by P. M. Orville of Geological Laboratory (now at Cornell University). The water recovered from a run of 11 s' duration in the presence of argon and within the stability field of analcite had a total sodium content of less than 25 ppm. It is possible that some dissolved material could have been precipitated by the sudden drop in pressure attendant upon making the analysis, but it is probable that the amount of dissolved material was not large enough to affect seriously the activity of the water in the mixture. Even if were present in large amounts, the amounts would probably be about the same as those in experiments performed with essentially pure water. It is concluded that the phase studies made in the presence of mixtures of water and argon are directly comparable to those made in the presence of initially pure water.

Comparison of experiments with theory. The results are presented in Table 11 and are shown

graphically in Figure 11. Each experiment indicated represents at least one run, and several of those at 2000 bars represent two or three runs. All but one of the indicated runs also represent reversal of the reaction, so that equilibrium is well demonstrated. The sizes of the plotted points indicate the estimated uncertainty in their positions.

The curved phase boundary in Figure 11 has been calculated from equation 11. A sample calculation follows, for the reaction written as



At 500°C the equilibrium pressure of the reaction in the presence of essentially pure water is 635 bars (Fig. 2). The boundary conditions for the calculation are therefore

$$T = 500^\circ\text{C}$$

$$P_{\text{H}_2\text{O}} = P_s = 635 \pm 50 \text{ bars}$$

$$\Delta V_s = -38.5 \text{ cm}^3$$

$$\Delta V_{\text{H}_2\text{O}} = 99.6 \text{ cm}^3$$

$$\left(\frac{\partial V_{\text{H}_2\text{O}}}{\partial P_{\text{H}_2\text{O}}} \right)_T = -0.20 \text{ cm}^3 \text{ bar}^{-1}$$

TABLE 11. Results of Experiments in the System $\text{NaAlSi}_3\text{O}_8\text{-H}_2\text{O-Argon}$

Pressure, bars	Temperature, °C	$x_{\text{H}_2\text{O}}$	$P_{\text{H}_2\text{O}}$, bars	Starting Materials	Duration, hours	Product	Reaction Product %
1965	501	0.525	1475	Anc + Ab + Ne	138	Anc + (Ab + Ne)	30
1960	500	.465	1415	Anc + Ab + Ne	188	Anc + Ab + Ne	0
2000	500	.418	1406	Anc + (Ab + Ne)	258	Ab + Ne + (Anc)	15
2000	500	.212	1120	Anc + (Ab + Ne)	167	Ab + Ne	100
1993	500	.310	1280	ox	168	Ab + Ne	100
2000	500	.393	1375	ox	192	Ab + Ne	100
2000	501	.354	1330	Anc + (Ab + Ne)	166	Ab + Ne + (Anc)	70
1506	500	.631	1155	Anc + Ab + Ne	142	Anc + (Ab + Ne)	30
1494	500	.460	992	Anc + (Ab + Ne)	264	Ab + Ne	100
1506	500	.503	1060	Anc + (Ab + Ne)	263	Ab + Ne + (Anc)	50
1100	500	.557	710	Anc + (Ab + Ne)	264	Ab + Ne	65
1110	499	.674	810	Anc + (Ab + Ne)	192	Ab + Ne	70
1050	500	.839	900	Ab + Ne + Anc	264	Anc + (Ab + Ne)	40
2010	500	.601	1575	Ab + Ne + (Anc)	140	Anc + (Ab + Ne)	85

Substitution of these values in equation 11 gives for the equation of the 500°C isotherm

$$P_{\text{H}_2\text{O}} = 1.5(10^{-4})P_s^2 + 0.196P_s + 450$$

Solving for $P_{\text{H}_2\text{O}}$ at $P_s = 2000$ bars gives

$$P_{\text{H}_2\text{O}} = 1442 \text{ bars, at } 500^\circ\text{C}$$

It can be seen from Figure 11 that the theoretical predictions and the experimental results are in excellent agreement over the range of pressures and at the temperature investigated.

Some uncertainty is associated with the above calculation. This uncertainty stems from three sources. The water content of the analcite may not be exactly as assumed. If the analcite were assumed to have a water/soda ratio of 2.00 instead of 1.87, the reaction would result in the evolution of more water than was used for the calculation shown. The equation for the isotherm under this assumption turns out to be

$$P_{\text{H}_2\text{O}} = 1.32(10^{-4})P_s^2 + 0.195P_s + 458$$

giving

$$P_{\text{H}_2\text{O}} = 1376 \text{ bars at } P_s = 2000 \text{ bars}$$

The uncertainty in the composition of the analcite therefore introduces an uncertainty in the calculated equilibrium pressure of 66 bars.⁴ This uncertainty is not very different from the

⁴ The equations for the reaction may be simplified by ignoring solid solution in analcite and nepheline. This leaves the final calculated P_s practically unchanged, owing to compensating errors.

other principal source of uncertainty, the pressure value of the equilibrium pressure in the presence of essentially pure water, which is known only to ± 50 bars. The third source of uncertainty is the assumption that the rate of change of slope of the isotherm is constant. Inclusion of this third order effect changes the calculated equilibrium pressure at $P_s = 2000$ bars by only 6 bars, and has therefore been neglected. In summary, most of the uncertainty in the calculated equilibrium pressure arises out of the uncertainty in the equilibrium pressure of the reaction at 500°C in the presence of essentially pure water, and it is this that forms the basis for drawing the shaded band of uncertainty around the calculated curve in Figure 11.

A test of the internal consistency of the data and the calculations can be made by making the necessary substitutions in equations 3, 19, 20, 21, and calculating the isobar along which $P_{\text{H}_2\text{O}} = 1442$ bars. Terms involving the thermal expansions and compressibilities of the solid phases are about 3 orders of magnitude less than the corresponding quantities for the fluid phases and so they may be safely neglected. Thus

$$\left(\frac{\partial^2 P_s}{\partial T^2}\right)_{P_s} = \frac{\Delta C_p}{T \Delta V_s}$$

where $\Delta C_p/T$ is given by equation 21. If

$$\frac{\Delta C_p}{T} \approx \text{const.},$$

$$P_s = \frac{1}{2}C_1T^2 + C_2T + C_3$$

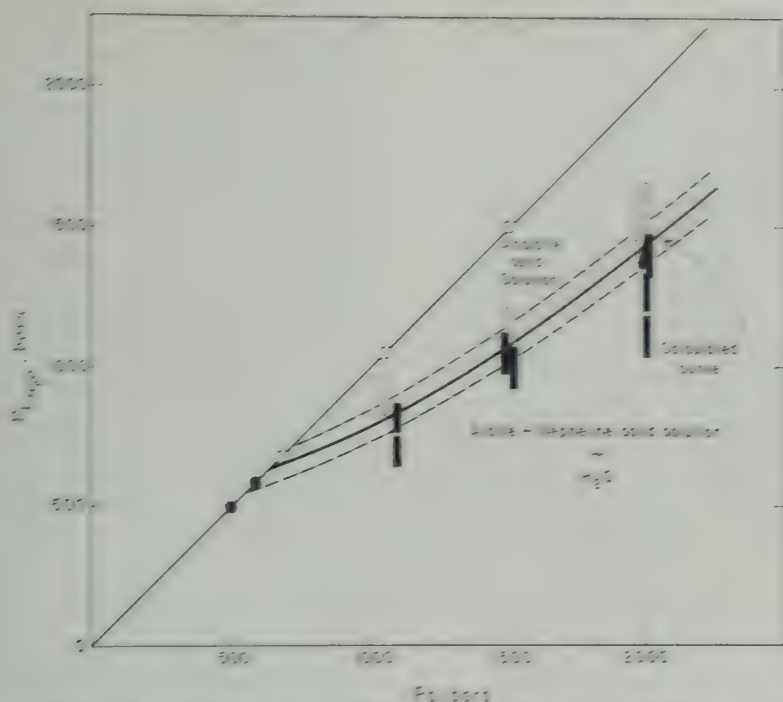


Fig. 11. Stability of ice with the total pressure, from the equilibrium position of ice at 50°C .

Figure 2, at $P = 1445$, $T = 54.5^\circ\text{C}$:

$$\left(\frac{\partial^2 P_E}{\partial T^2}\right)_{P_E} = 0$$

The equation of the isobar is therefore a straight line of the form

$$P_E = C_2 T + C_1$$

Letting $t = T^\circ\text{C} - 450$, the isobar becomes

$$P_E = C_2 t + C_1'$$

$$C_2 = \left(\frac{\partial P_E}{\partial T}\right)_{P_E} = -13.6 \text{ bar deg}^{-1}$$

and

$$P_E = -13.6t + 2735$$

At 50°C , $t = 50$, and $P_E = 2054$ bars. This 54-bar difference from the 2000 bars assumed when calculating the isotherm represents an 'error of closure' that is about the same magnitude as the uncertainty in calculating the isotherm, and therefore represents a satisfactory check on the equilibrium positions of the curves and on the calculations.

Mention has been made that diagrams showing the temperature, total pressure, and equilib-

$$\frac{dP}{dT} = 21 \pm 3 \text{ bar deg}^{-1}$$

$$\frac{d^2P}{dT^2} = 0.20 \pm 0.03 \text{ bar deg}^{-2}$$

$$\Delta V_E = -14.5 \pm 1 \text{ cm}^3$$

$$V_{\text{NaNO}_3} = 61.6 \text{ cm}^3$$

Bar and Kennedy, 1959.

$$\Delta V_E + V_{\text{NaNO}_3} = 25 \pm 1 \text{ cm}^3$$

$$\left(\frac{\partial V_E}{\partial T}\right)_{P_E} = 0.147 \text{ cm}^3 \text{ deg}^{-1}$$

$$\left(\frac{\partial V_E}{\partial P}\right)_{P_E} = -1.45(10^{-4}) \text{ cm}^3 \text{ bar}^{-1}$$

Bar and Kennedy, 1959.

and

$$\frac{\partial V_E}{\partial T} = -0.8 \pm 0.3 \text{ bar}^{-1} \text{ deg}^{-1}$$

and $\Delta C_p/T$ is non-negligible from zero

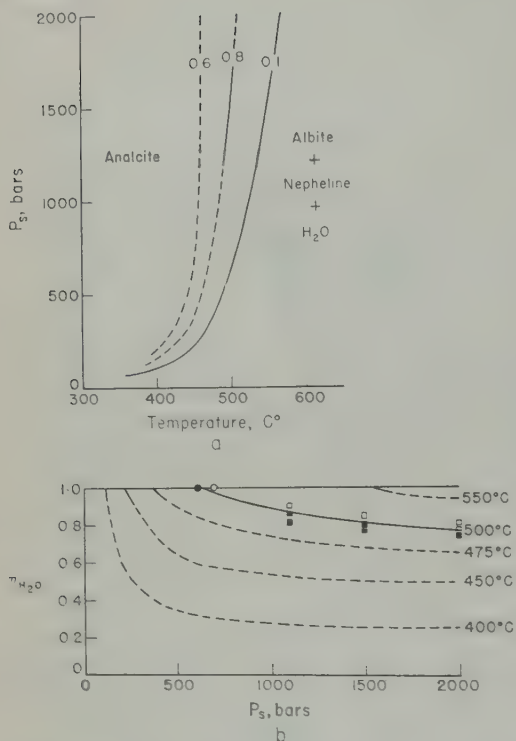


Fig. 12. Stability of analcrite as a function of pressure, temperature, and humidity of water.

rium pressure are not the only way to represent data like those shown in Figure 11. Reference to equation 26 indicates that the equilibrium pressure of water might be translated into the 'humidity' (Humidity equals mole fraction in ideal mixtures) or fugacity ratio η_{H_2O} , and that a different diagram might be drawn. This transformation has been made, and the resulting diagram is shown in Figure 12. This kind of diagram has the advantage of being meaningful throughout its whole three-dimensional volume, whereas the equilibrium pressure diagram is not. Equilibrium pressures greater than the total pressure are impossible, so that the region that lies above the plane where $P_x = P_z$ is not occupied in the equilibrium pressure diagram. However, this diagram has the advantage of being more closely related to a measurable pressure than the 'humidity,' and in the case of rock systems that are open to water the equilibrium pressure may be thought of as the water pressure in the surrounding rocks that act as a water reservoir. The choice between the use of either humidity or equilibrium pressure must in

the end be based on intended application, and each is equally valid for describing the same phenomena.

APPLICATIONS TO GEOLOGY

The environments possible in both rocks and laboratory experiments may be divided into two broad classes: those in which the equilibrium pressure of the reacting volatile component is essentially equal to the pressure on the solid phases, and those in which it is less. This second class may be realized in several ways. The first way, and possibly the most important, is realized in the presence of a fluid phase that has more than one component. The second way is realized in the presence of a pore fluid that supports only part of the rock load, the balance being supported by the strength of the mineral grains. Equilibrium is attained only locally, and the minerals are at least partly under nonhydrostatic stress. The third way is realized in the presence of a volatile component that has its chemical potential controlled by some agent outside the system in question.

Pore Fluids with More Than One Component

Mineral equilibria taking place in nature probably almost never occur in the presence of a pure one-compound fluid phase. This is particularly true during the metamorphism of weakly impure calcareous sediments. Consequently the temperature represented by a particular isograd for a mineral reaction will depend not only on the total pressure but also on the mole fraction of the reacting volatile component in the fluid phase. A first approximation to the effect of diluting the reacting volatile with other components can be made on the assumption of ideal mixing. This effect is modified by the effect of keeping the total pressure constant while reducing the equilibrium pressure through dilution, as discussed in the early part of this paper. These remarks, although not profound, are made to emphasize the fact that isograds that are defined by reactions involving volatile components should not be regarded as isotherms even within the confines of a single map area, unless there is some evidence that the composition of the pore fluid was approximately constant over the area. This enhances, rather than reduces, the importance of mapping isograds. Displacement of isograds from their usual geographic relation

with other isograds and sharp changes in an isograd can tell us more in principle, of the composition of the pore fluids and the temperature environment than isograds appear in their normal positions.

Equilibrium near Pores, with Nonhydrostatic Stress

Some rocks having their pores filled with a fluid that sustains a lower pressure than the solid grains are in a state of mechanical disequilibrium. Grains around such pores will tend to dissolve and recrystallize so as to fill the pores, distribute the load hydrostatically, reduce the volume of the rock mass and tending to relieve the nonhydrostatic stress on the minerals around the pore. The net tendency will be to eliminate pores and the nonhydrostatic stress simultaneously.

The elimination of the nonhydrostatic stress and the pores is controlled by a series of competing rate processes, which may be conveniently divided into two groups—the diffusion of reaction products to and from the site of the pore and the rate of the reaction itself. If reaction products are able to enter or leave the site of the reaction faster than they are consumed or produced, then the controlling step will be reaction rate. In prograde metamorphism, where the evolution of a volatile component is important, this would result finally in a system without pores, consisting of a mineral assemblage appropriate to the temperature and pressure of the rock, and to the equilibrium pressure of the volatile component in the surrounding rocks. Such a system would be considered open to the volatile component.

If the reaction products cannot escape as fast as they are produced, owing to a low connected porosity the recrystallization will proceed at a gradually decreasing rate, with the pore fluid gradually changing composition and taking an increasingly large proportion of the rock load. The process should come to complete equilibrium, the final rock would be porous, the pores would be filled with fluid at a pressure equal to the pressure on the mineral grains, and the system would be under simple hydrostatic stress. This equilibrium configuration corresponds to the first example discussed in which the principal complicating factor is the composition of the pore fluid.

The natural occurrence of pores of the kind described is widespread, but the approach to equilibrium may not be. Deeply buried sandy rocks commonly preserve some of their porosity, and do not always have their pore fluids at the total pressure of the rock. There is no doubt that these rocks are mechanically unstable, and their preservation merely points to the slow rate of reaction under such conditions.

Rocks pierced by drill holes in the Taringatura hot spring area of New Zealand show some evidence that the load pressure has been in excess of the pressure on the pore fluid. Ellis and Fyfe [in *Coombs, Ellis, Fyfe and Taylor, 1959, p. 83*] describe zones of relief heulandite in the higher temperature laumontite zone, and note that the heulandite occurs in rocks of low permeability, while the laumontite occurs in coarser grained and more porous rocks. They postulate that in the less permeable rocks the water was at the same pressure as the mineral grains, whereas in the more permeable rocks the water was able to diffuse away from the site of reaction to a lower pressure region, giving rise to less hydrated (higher temperature) minerals such as laumontite.

Rocks without Pores

Condensed systems contain no observable pores, but they may contain a component that is potentially volatile and that has an equilibrium pressure less than the pressure on the solid phases. This equilibrium pressure has been called a fictive pressure [Foder, 1955], to emphasize that it is not exerted by the pure component present as a phase. Thompson [1959] has shown that no such component, whose chemical potential is externally controlled, may have the composition of any of the phases present (inactive components). The phase rule for open systems [Thompson, 1959; Korzhinskii, 1959] assumes the possibility of condensed systems being open to some components whose chemical potentials are externally controlled. A system can come to equilibrium with such a component only if the component has access to all parts of the system. In the absence of pores such access must be gained through other channels, such as grain boundaries or lattice imperfections. It might seem at first that admitting the presence of channels of any sort admits the existence of the component as a separate phase within the channel,

however small the channel may be, and that to maintain otherwise is merely to quibble about the size of the opening one must recognize before admitting the presence of a phase. However, this apparent indefiniteness is not real if the existence of a pure one-component phase is regarded from the standpoint of the free energy of the phase. If a component in the system has a Gibbs' free energy equal to the free energy of the pure component taken at the pressure and temperature of the system, then the phase rule requires that the component be present as a discrete phase. If the component in the system has a free energy less than this value, it cannot appear as a phase, because any such phase would change spontaneously to the lower free-energy condition of the component in the system. It is clear, therefore, that no component to which the system is open can contribute an extra phase to the system if the partial molar free energy of the component is less than the free energy of the pure component taken at the pressure and temperature of the system. The physical location of the mobile component that is not present as a phase is most probably to be found along grain boundaries and lattice imperfections. Grain boundaries are not normally considered to represent separate phases in polycrystalline aggregates because they make a wholly negligible contribution to the total free energy of the system. Additions of small amounts of the mobile component to the grain boundaries may change the composition of these regions but need not destroy their continuity. More detailed expressions of the phase rule that include the intergranular regions as phases still need not include the mobile components as contributors to the number of necessary phases.

The physical possibility of a chemical species migrating along grain boundaries without adding a phase to the assemblage is attested to by experiments made by metallurgists on diffusion of one metal into another. In such experiments the migrating component is often detected only by autoradiography, and extra phases are not observed. Such diffusion is fastest along grain boundaries and slowest through perfect crystals. Extrapolation of these observations to geological processes is not without pitfalls, but it seems reasonable to suppose that a volatile component could move through and come to equilibrium with a condensed rock system.

Preliminary experiments, reported elsewhere [Greenwood, 1960, p. 63], support this point of view. Water was shown to be capable of diffusing through a fine-grained aggregate of brucite and of reacting with periclase to form malachite. This diffusion occurred at a temperature of 500°C, with the brucite under an external confining pressure several times the pressure of the pure water that was allowed to diffuse. The water was diffusing down the chemical potential gradient of water, but *up* the total pressure gradient. Thin sections of the final product show no pores, only a compact pellet. The pressure of the water in the reservoir was too low to force pores to open in the brucite aggregate, even temporarily, and it seems reasonable to suppose that the movement of water took place along grain boundaries, and possibly in pores through lattice imperfections. These experiments, although preliminary, provide evidence that water, at least, can diffuse through a system consisting only of condensed phases and could therefore equilibrate with these phases.

'Water-deficient' systems [Yoder, 1952] have been described as those with insufficient water to convert all the anhydrous phases present to the corresponding hydrous phases that are stable at that pressure and temperature. Such systems are distinguished by the coexistence of the solid reactants and solid products of a reaction involving water, and the absence of water as a phase. Such a phase assemblage defines the equilibrium pressure of water in the rock at an arbitrary temperature and pressure of the solid phases, and consequently such assemblages can be treated in the same frame of reference as solid condensed open systems, using the variables T , P_s and P_w .

Demonstration of the physical possibility of the equilibration of condensed rock systems with external reservoirs of water does not mean that all rocks can or have come to equilibrium with water in such a fashion. The observation is common that the strongly metamorphosed equivalents of water-rich sediments contain less water than the original sediments. These rocks have evidently been open to the loss of water, as in the cases where retrogressive metamorphism has been accompanied by the development of hydrous minerals, also open to the addition of water. The fact that the analyses of the mineral assemblages of pelitic schists presented

Thompson [1957] and by Zen [1960] are internally consistent indicates that the concept of closed systems open to water is probably valid in many instances. Some rocks afford good evidence of having been closed to water, however, depending on the mineral assemblage present. The equilibrium pressure of water is either defined or defined by mineralogy, rather than the equilibrium pressure of water in the surrounding environment.

An excellent example of such a closed confined system is given by Zen [1961] from the pyrophyllite deposits of North Carolina. He points out the evidently stable coexistence of kaolinite, pyrophyllite, and quartz, which are chemically related to one another by the reaction $\text{kaolinite} + \text{quartz} = \text{pyrophyllite} + \text{water}$. This assemblage is sufficient to define the $P_{\text{H}_2\text{O}}$ of the rock at any particular total pressure and temperature. If the rock were open to water, the rock would have had to be at exactly the correct temperature for a particular total pressure for the assemblage to be stable. This probable coincidence of conditions suggests that the rock was closed to water. Zen points out further that the rock was closed to water on a microscopic scale. The absence of water as a phase in a rock such as this does not mean that the equilibrium pressure of water was either higher or equal to the total rock pressure. The definition of equilibrium pressure in terms of chemical potential indicates that the pressure must be finite and less than the total pressure of the mineral grains. In a closed system, water in such a condition cannot be present as a pure phase. It is probably present partly as water of hydration in the kaolinite and pyrophyllite, and partly as an adsorbed film on the grain boundaries of the mineral. Such adsorption would be expected to lower the free energy of the water, presumably enough to let it be at equilibrium with the phases that are reported. Other examples of systems closed to water might be added, such as the adjacent bands of amphibolite and gabbro xenite granulite reported by Eskola [1952].

APPLICATIONS TO EXPERIMENTAL WORK

The good agreement between the theory and experiments, illustrated in Figure 6, suggests a direct application of the theory to other experimental data. Phase-equilibrium studies on solid systems in the presence of contaminated fluids can

be used to calculate the equilibrium pressure of the component in the fluid that is taking part in the reaction. The displacement of the equilibrium as a function of fluid composition, pressure, and temperature would have to be observed experimentally, and a diagram similar to Figure 1b constructed by application of the equations to account for the effect of total pressure on the equilibrium. Combination of the calculated dependence of P_B and T on total pressure with the measured dependence on fluid composition, pressure, and temperature would permit calculation of the equilibrium pressure of the reacting component in the fluid phase. This method could furnish information only over a curved surface in the P - T - X space, but combination with several equilibria observed to be displaced by the same contaminated fluids at different pressures and temperatures would provide a good reconnaissance of the whole system.

An application of this method to the system $\text{NaOH--H}_2\text{O}$ using the stability of brucite, $\text{Mg}(\text{OH})_2$, as the indicating equilibrium, is described by Barnes and Ernst [1961].

Acknowledgments. I was supported during the course of this research by the Porter Ogden Jacobus fellowship and by the Siscoe fellowship, both of Princeton University, and by a predoctoral fellowship of the Carnegie Institution of Washington, at the Geophysical Laboratory. Without the help of these institutions the work could not have been done, and I am grateful to them.

Many people have actively helped the progress of this work by discussion and contribution of their time and experience. Special thanks are due F. R. Boyd and S. P. Clark, Jr., whose critical comments were invaluable. Helpful discussion and criticism have been offered by Professors H. H. Hess and H. D. Holland of Princeton University; Professor J. B. Thompson, of Harvard University; H. S. Yoder, H. L. Barnes, and W. G. Ernst, of Geophysical Laboratory; D. R. Wones, of the U. S. Geological Survey; and R. I. Harker, of Johns-Manville Research Center.

REFERENCES

- Barnes, H. L., and W. G. Ernst, Ideality and ionization in hydrothermal fluids: the system $\text{MgO--H}_2\text{O--NaOH}$, *Am. J. Sci.*, in press, 1961.
- Bennett, C. O., Fugacities of gases in mixtures, *Chem. Eng. Progr. Symposium Ser.*, no. 49 (7), 45-52, 1953.
- Coombs, D. S., A. J. Ellis, W. S. Fyfe, and A. M. Taylor, The zeolite facies with comments on the interpretation of hydrothermal syntheses, *Geochim. et Cosmochim. Acta*, 17, 53-107, 1959.
- Danielsson, A., Das Calcit-Wollastonitgleichge-

- wichte, *Geochim. et Cosmochim. Acta*, **1**, 55-59, 1950.
- Donnay, G., J. F. Schairer, and J. D. H. Donnay, Nepheline solid solutions, *Mineral. Mag.*, **32**, 93-109, 1959.
- Eskola, P., On the granulites of Lapland, *Am. J. Sci., Bowen Vol.*, 133-171, 1952.
- Fyfe, W. S., F. J. Turner, and J. Verhoogen, Metamorphic reactions and metamorphic facies, *Geol. Soc. Am. Mem.*, **73**, 35-40, 1958.
- Gillespie, L. J., Expansion of gases in mixing, especially at very low pressures, 1. Its relation to the empirical calculation of the fugacities in gaseous mixtures, *Phys. Rev.*, **34**, 352-356, 1929.
- Greenwood, H. J., Water pressure and total pressure in metamorphic rocks, *Carnegie Inst. Washington Year Book*, **59**, 58-63, 1960.
- Hamilton, D. L., Nephelines as crystallization temperature indicators, *J. Geol.*, **69**, 321-329, 1961.
- Harker, R. I., The system $\text{MgO}-\text{CO}_2$ -argon, and the effect of inert pressure on certain types of hydrothermal reaction, *Am. J. Sci.*, **256**, 128-138, 1958.
- Hilsenrath, J., and others, Tables of thermal properties of gases, *Nat. Bur. Standards (U. S.), Circ.* **564**, 1955.
- Holser, W. T., Fugacity of water at high temperatures and pressures, *J. Phys. Chem.*, **58**, 316-317, 1954.
- Holser, W. T., and G. C. Kennedy, Properties of water, 5. Pressure-volume-temperature relations of water in the range $400^\circ\text{C}-1000^\circ\text{C}$ and 100-1400 bars, *Am. J. Sci.*, **257**, 71-77, 1959.
- Kennedy, G. C., Pressure-volume-temperature relations in water at elevated temperatures and pressures, *Am. J. Sci.*, **248**, 540-564, 1950.
- Korzhinskii, D. S., *Physicochemical Basis of the Analysis of the Paragenesis of Minerals*, Consultants' Bureau, Inc., New York, 142 pp., 1959 (translated from the Russian).
- Michels, A., H. Wijker, and H. K. Wijker, Isotherms of argon between 0°C and 150°C and pressures up to 2900 atmospheres, *Physica*, **627-633**, 1949.
- Saha, P., Geochemical and X-ray investigations of natural and synthetic analcite, *Am. Mineralogist*, **44**, 300-313, 1959.
- Saha, P., System nepheline albite, *Am. Mineralogist*, **46**, 859-884, 1961.
- Sand, L. B., R. Roy, and E. F. Osborn, Stability relations of some minerals in the $\text{Na}_2\text{O}-\text{Al}_2\text{O}_3-\text{SiO}_2-\text{H}_2\text{O}$ system, *Econ. Geol.*, **52**, 169-179, 1957.
- Schairer, J. F., and N. L. Bowen, The system $\text{K}_2\text{O}-\text{Al}_2\text{O}_3-\text{SiO}_2$, *Am. J. Sci.*, **253**, 681-746, 1955.
- Smith, J. V., The powder patterns and lattice parameters of plagioclase feldspars, 1. The sodic rich plagioclases, *Mineral. Mag.*, **31**, 47-68, 1954.
- Thompson, J. B., Jr., The thermodynamic basis for the mineral facies concept, *Am. J. Sci.*, **260**, 65-103, 1955.
- Thompson, J. B., Jr., The graphical analysis of mineral assemblages in pelitic schists, *Am. Mineralogist*, **42**, 842-858, 1957.
- Thompson, J. B., Jr., Local equilibrium in metamorphic processes, in *Researches in Geochemistry*, edited by P. H. Abelson, John Wiley & Sons, New York, pp. 427-457, 1959.
- Yoder, H. S., Jr., The $\text{MgO}-\text{Al}_2\text{O}_3-\text{SiO}_2-\text{H}_2\text{O}$ system and the related metamorphic facies, *Am. J. Sci., Bowen Vol.*, 569-627, 1952.
- Yoder, H. S., Jr., Analcite, *Carnegie Inst. Washington Year Book* **53**, 121-122, 1954.
- Yoder, H. S., Jr., The role of water in metamorphism, in *The Crust of the Earth*, *Geol. Soc. Am. Spec. Paper* **62**, 505-524, 1955.
- Zen, E-an, Metamorphism of Lower Paleozoic rocks in the vicinity of the Taconic range, west-central Vermont, *Am. Mineralogist*, **45**, 129-175, 1960.
- Zen, E-an, Mineralogy and petrology of the system $\text{Al}_2\text{O}_3-\text{SiO}_2-\text{H}_2\text{O}$ in some pyrophyllite deposits of North Carolina, *Am. Mineralogist*, **46**, 52-66, 1961.

(Manuscript received August 3, 1961.)

Geomagnetic and Solar Data

J. VIRGINIA LINCOLN

CRPL, National Bureau of Standards
Boulder, Colorado

INTERNATIONAL DATA ON MAGNETIC DISTURBANCES

This report continues the series which has appeared regularly in this journal since Volume 65, no. 3, 295 (1949). Please refer to that first report for an explanation of the data given, and to Volume 59, no. 3, 423 (1954) for the definition of *Ap*.

Note: Additional and final Geomagnetic and Solar Data appears in due course in the following international publications: "*Quarterly Bulletin on Solar Activity*," International Astronomical Union, c/o Eidgen. Steinwarte, Zurich, Switzerland; "*IAGA Bulletins, Geomagnetic Indices and C₁*," by J. Bartels, A. Romafña and J. Adkamp, International Union of Geodesy and Geophysics, Association of Geomagnetism and Aeronomy, c/o V. Laursen, Meteorologisk Institut, Charlottenlund, Denmark.

SUDDEN COMMENCEMENTS AND SOLAR-FLARE EFFECTS, FIRST QUARTER, 1961

Preliminary Report of Sudden Commencements

Records given by ten or more stations are in *italics*. Times are mean values obtained from normal magnetograms.

Sudden commencements followed by a magnetic storm or a period of storminess (s.s.c.)

1961 January 07d 20h 49m: Me Wn? Ma Db
Te Gn (si: El Tn).—08d 16h 17m: forty-one
c: 26; si: 14; bs: 1).—18d 02h 00m: MB Va.—
12h 50m: Ma Db.—30d 21h 22m: Tr So
(TI).

1961 February 03d 09h 08m: forty-one (ssc:
si: 1).—04d 13h 31m: forty-two (ssc: 39;
3).—04d 18h 29m: thirty-one (ssc: 21; si: 10).—
01h 06m: forty-four (ssc: 33; si: 10; pg: 1).—
02h 53m: forty-four (ssc: 39; si: 5).—

13d 07h 39m: Sw VI (pt: To).—16d 00h 26m:
Sw Te.—16d 00h 43m: forty-six (ssc: 44; si: 2).—
16d 05h 36m: fourteen (ssc: 13; si: 1).

1961 March 09d 13h 27m: fifty-five (ssc: 53;
si: 2).—19d 10h 26m: SF Mc?—27d 15h 03m:
forty-nine.—31d 15h 12m: forty-nine (ssc: 39;
si: 10).—31d 23h 24m: Me Gn.

Sudden impulses found in the magnetograms (s.i.)

1961 January 05d 08h 29m: Tr Ks (pt: Wn?
Ci).—15d 06h 12m: VI IK MB.—18d 06h 20m:
IK Ka (ssc: MB; bs: PM).—24d 07h 26m:
El Hr (ssc: Gn; bs: To; pt: TI).—25d 22h 26m:
Tr (ssc: So; sfc: Me?).

1961 February 03d 10h 21m: Pr Te.—03d
19h 08m: Te MB.—05d 02h 48m: eleven.—
06d 06h 19m: Qu Ho.—06d 06h 42m: SF Ka Ho
Te Bi PM (ssc: Mb Ky MB Mc).—16d 10h 05m:
CF Qu MB.—17d 14h 00m: sixteen (si: 14; ssc:
2).—17d 17h 20m: Mb Ka Ky PM El (ssc: Ho;
pt: Cm).—17d 17h 33m: VI Pr Eb Ho MB PM.—
17d 20h 25m: twelve (si: 10; ssc: 2).

1961 March 09d 21h 45m: MB (ssc: Mc).—
09d 22h 39m: Me Te.—10d 11h 03m: nineteen
(si: 18; pt: 1).—18d 20h 17m: Me (ssc: Qu).—
21d 11h 30m: IK (sfc: MB?).—28d 01h 56m:
twenty-two (si: 12; ssc: 9; bs: 1).

Preliminary report on Solar-Flare Effects (s.f.e.)

Effects confirmed by ionospheric or solar
observations are in *italics*.

1961 January 01d 03h 56m: To.—01d 13h 21m:
IK.—03d 19h 00m: Hu (si: Me).—04d 02h 07m:
Mb Ka Ky Gn (si: Gu PM).—05d 13h 52m:
Hu.—07d 21h 46m: Te.—08d 14h 03m: Te (ssc:
Sw).—15d 19h 10m: Si.—25d 14h 18m: Te.—
26d 09h 50m: MB.—27d 23h 18m: Si (si: Me).—
29d 15h 31m: Hu.—30d 06h 30m: Gn.—30d 11h
04m: IK.—30d 14h 22m: MB Hr.—30d 20h 04m:
Tu.

1961 February 01d 09h 04m—09h 32m: TI.—
04d 17h 03m: Hu.—06d 18h 46m: Te.—06d 21h
05m: Te.—08d 09h 23m: IK.—17d 13h 10m:

3948

J. VIRGINIA LINCOLN

Te.—19d 21h 16m: Te.—20d 13h 20m: Te.—21d 21h 27m: Te.—27d 15h 06m: VI Ma? Tu MB Hu (si: Es HI).—28d 14h 39m: VI Tu MB.

1961 March 01d 14h 22m: Hu (pt: Me).—11d 16h 55m: Si (si: Me).—12d 13h 39m: Fr.—14d 13h 10m: Te.—14d 21h 27m: Te.—15d 15h 12m: Te (si: Me).—16d 09h 42m–09h 48m: Eb.—16d 17h 57m: Hu.—26d 10h 22m–11h 28m: Db

Eb Tn.—26d 10h 40m: Es.—28d 13h 49m: Hu.—29d 04h 12m: Ho.—29d 15h 58m: Hu.

COMMITTEE ON RAPID VARIATIONS AND EARTH CURRENTS

A. ROMANA, Chairman, Observatorio del Ebro, Tortosa, Spain

PRINCIPAL MAGNETIC STORMS.

(Advance knowledge of the character of the records at some observatories as regards disturbances)

Observatory (Observer-in-Charge)	Greenwich date	Storm-time		Sudden commencement			C-figure, degree of activity ⁴	Maximal activity on <i>K</i> -scale 0 to 9			Ranges			
		GMT of begin.	GMT of ending ¹	Type ²	Amplitudes ³			Gr. day	Gr. 3-hr. period	<i>K</i> -index	<i>D</i>	<i>H</i>	<i>Z</i>	
					<i>D</i>	<i>H</i>								<i>Z</i>
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Hermanus (A. M. v. Wijk)	1961	<i>h m</i>	<i>d h</i>			<i>γ</i>	<i>γ</i>							
	Apr. 9	06 ..	11 19				m	9	4	5	19	108	7
	Apr. 13	14 50	15 17	s.c.	+1.3	+9	+8	ms	5	1	7	37	137	18
	May 6	14 ..	6 16	Minor disturbance	m	6	2.3	5
	May 24	23 ..	26 04	m	25	2	5	25	104	6
	June 2	07 ..	2 19	Minor disturbance	m	2	6	5
	June 6	17 ..	7 00	Minor disturbance	m	6	6	5
	June 20	16 ..	23 04	ms	21	8	6	26	121	12
								22	1	6	
	June 29	00 ..	29 12	ms	29	2	6	13	69	6

¹Approximate time of ending of storm construed as the time of cessation of reasonably marked disturbance movements in the traces; more specifically, when the *K*-index measure diminished to 2 or less for a reasonable period.

²s.c. = sudden commencement; s.c.* = small initial impulse followed by main impulse (the amplitude in this case is that of the main impulse only, neglecting the initial brief pulse); ... = gradual commencement.

³Signs of amplitudes of *D* and *Z* taken algebraically; *D* reckoned positive if towards the east and *Z* reckoned positive if vertically downwards.

⁴Storm described by three degrees of activity: *m* for moderate (when *K*-index as great as 5); *ms* for moderately severe (when *K* = 6 or 7); *s* for severe (when *K* = 8 or 9).

SELECTED GEOMAGNETIC AND SOLAR DATA

Kp, Ci, Cp, Ap, K_F, Rz and Selected Days

July 1961

Day ¹	Three-hour range indices <i>Kp</i> ²									Prel. ³ <i>Ci</i>	<i>Cp</i> ⁴	<i>Ap</i> ⁵	3-hr. range indices <i>K_F</i> ⁶		Prov. ⁷ <i>Rz</i>
	1	2	3	4	5	6	7	8	Sum				Values	Sum	
q	3-	3-	3-	3o	2o	3-	0+	1o	17o	0.4	0.5	10	3332 1211	16	60
q	1+	2+	3-	2+	2-	3+	4-	3o	20+	0.7	0.7	12	1332 2333	20	65
q	2-	2o	3+	4o	4+	4o	2+	1+	23o	0.9	0.9	16	2233 3322	20	68
q	2+	1+	2-	2+	3o	3-	3o	5+	22-	1.0	0.9	16	3222 3335	23	63
D	6o	6+	4o	4o	5o	5-	3o	4o	37o	1.4	1.5	45	4634 4434	32	44
q	3o	4o	3-	2+	2o	5-	3-	2+	24-	0.9	0.9	16	4522 3333	25	50
q	4o	3-	3+	3o	3+	3-	2-	1-	21+	0.8	0.8	14	4341 3321	21	51
q	2-	2o	3-	3-	3-	2+	2-	3-	18+	0.5	0.5	10	2333 2233	21	60
q	2+	2+	4-	2o	1+	3-	1o	1-	16o	0.5	0.5	9	3342 1211	17	65
q	3-	4o	4-	3-	3-	3+	2-	2o	23-	0.8	0.8	14	2433 2222	20	73
Q	3o	2o	1+	1+	1+	3-	3-	2+	17-	0.4	0.5	9	3222 1222	16	85
Q	2-	1o	3-	1o	1o	0+	1+	0+	9+	0.4	0.2	5	2134 1011	13	96
D	0+	1o	1-	6+	8o	8+	8-	6+	39-	1.8	1.9	102	1105 7755	31	86
D	5-	3o	7-	8+	7+	7-	6+	4o	47o	1.8	1.9	98	3356 5553	35	107
q	3o	5-	3o	3o	4+	2+	5-	5-	30-	1.2	1.2	25	3434 3144	26	100
q	6-	4o	3-	2+	3+	3+	4-	3-	28-	1.1	1.1	23	5421 3333	24	94
q	3o	3-	2+	4-	4+	2+	6+	6+	31o	1.4	1.4	36	3323 3265	27	92
D	5o	6+	6o	5o	8-	7o	6-	6+	49o	1.8	1.9	93	5554 6555	40	82
q	6o	5-	1+	0+	2-	1o	1o	1+	17+	0.8	1.0	18	5410 2102	15	86
q	2o	3o	3+	2o	2-	3o	4o	5+	24+	1.1	1.0	19	3332 2445	26	85
q	5o	6-	6-	3-	3+	4o	4o	3-	33o	1.3	1.4	35	4563 2333	29	85
q	3-	3-	4-	3o	3-	3-	1-	1o	19o	0.6	0.7	12	3332 2211	17	75
q	1o	1o	2-	3+	4-	5o	4-	3o	22+	0.9	0.9	17	1124 3333	20	81
q	2o	3+	3+	3+	3+	2+	1+	2+	21+	0.7	0.7	13	2444 2212	21	78
q	4o	3+	3+	3o	2+	2-	2o	2+	22o	0.7	0.8	14	3434 2123	22	63
q	4o	4-	2-	2o	2-	1o	5+	5+	25-	1.1	1.1	23	3413 1155	23	62
D	4o	4+	8+	8+	6o	5+	6+	7+	50o	1.9	1.9	114	4478 5566	45	53
q	4o	3o	3+	3+	2o	4-	4o	2+	26-	1.0	1.0	18	4443 2332	25	42
Q	2-	3+	2+	1+	2+	1+	2-	2o	16o	0.5	0.4	8	1421 2222	16	32
Q	2+	2-	3o	2-	1+	1o	2+	2o	15+	0.6	0.4	8	2232 1122	15	30
Q	1+	1-	1o	2-	3-	2-	2+	2-	13o	0.3	0.3	6	2111 2221	12	34
Means:										0.94	0.96	28			69.3
No. of days:										31	31	31			31

Notes:

¹ Five quiet days (Q), ten quiet days (Q or q), five disturbed days (D) selected by Committee on Characterization of Magnetic Disturbances, J. Veldkamp, Kon. Nederlandsch Meteorologisch Instituut, DeBilt, Holland.² Geomagnetic planetary three-hour-range indices *Kp* prepared by Committee on Characterization of Magnetic Disturbances, J. Bartels, Chairman, University, Göttingen, Germany.³ Preliminary magnetic character-figures, *Ci*, prepared by J. Veldkamp.⁴ Magnetic character-figures, *Cp*, prepared by J. Bartels.⁵ Average amplitudes *Ap* (unit 2 γ), prepared by J. Bartels.⁶ Fredericksburg three-hour-range indices *K* (*K*₉ = 500 γ); scale-values of variometers in γ /mm: *D* = 2.8; *E* = 2.5, *Z* = 3.0) prepared by Robert E. Gebhardt, Observer-in-Charge, Fredericksburg Magnetic Observatory, Corbin, Virginia.⁷ Provisional sunspot-numbers (dependent on observations at Zurich Observatory and its stations at Locarno and Arosa) prepared by M. Waldmeier, Swiss Federal Observatory, Zurich, Switzerland.

Letters to the Editor

The Primary Cosmic-Ray Electron Flux during a Forbush-Type Decrease

PETER MEYER AND ROCHUS VOGT

*Enrico Fermi Institute for Nuclear Studies
University of Chicago, Chicago, Illinois*

The discovery of primary cosmic-ray electrons in the vicinity of the earth [Meyer and Vogt, 1961; Earl, 1961] opens the question of the origin of these particles. There exist two obvious alternatives, namely (1) solar origin with subsequent storage in interplanetary space, and (2) galactic origin. In the second case the electrons would most likely be identical with the long postulated source of galactic radio noise. Their intensity and energy spectrum near the

earth would be modified by the modulation mechanisms that are known to affect the flux of protons arriving from the galaxy. This modification would be strongest during periods close to the maximum of solar activity.

It is important to make a decision between the two alternatives of solar and galactic origin. Although there exists no experimental evidence as yet leading to an unambiguous answer, we are able to report here on some results that bear on this question.

These results concern:

1. The production of electrons at the sun during the September 3, 1960, solar flare, which is known to have produced a large flux of high energy protons.
2. The short-term intensity variation of the primary electron flux during a Forbush-type decrease.

The results are based on measurements carried out at balloon altitudes over Fort Churchill, Manitoba, on August 22, September 8, and September 15, 1960. The equipment consisted of a scintillation counter telescope, designed to measure the initial energy loss and the range in lead of low energy primary particles. A cross section of the detector is shown in Figure 1. Full details of its properties have been published elsewhere [Meyer and Vogt, 1961; Vogt, 1962].

In Figure 2 the neutron monitor data of the Deep River station (kindly made available to us by Drs. H. Carmichael and T. Steljes) are shown for the period in which the balloon measurements were made. On September 3, a solar flare occurred that resulted in the emission of solar protons with energies up to several hundred Mev. These protons were studied through balloon and rocket measurements by Winckler, Bhavsar, Masley, and May, 1961, and by Davis, Fichtel, Guss, and Ogilvie, 1961. The energy spectrum of the stored flare particles has also

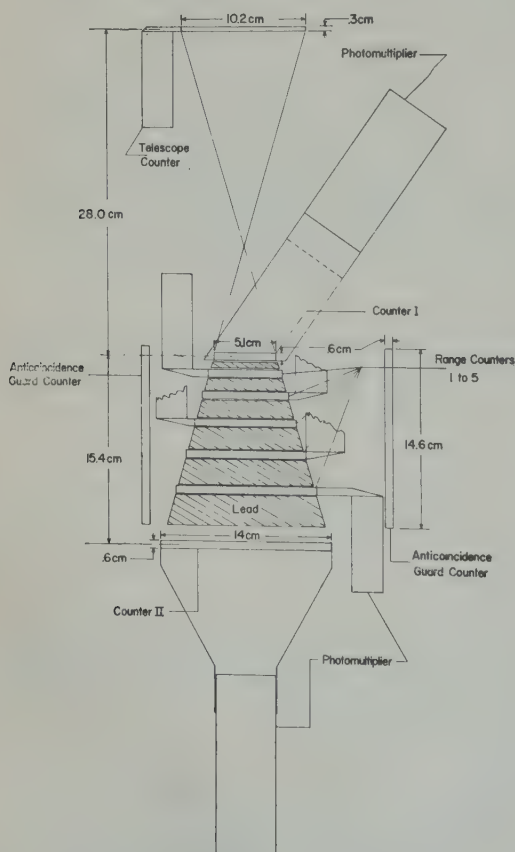


Fig. 1. Cross section of the detector system.

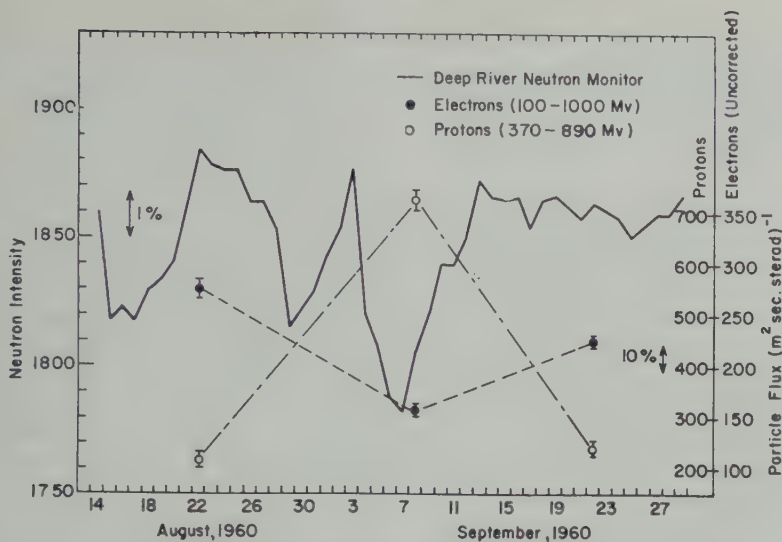


Fig. 2. Daily averages of Deep River neutron monitor data during period of balloon measurements. Also shown are the proton and uncorrected electron flux in similar rigidity intervals for the 3 days of measurement.

investigated with our apparatus on September 8, five days after their emission [Vogt, 1960], when the total cosmic-ray intensity had undergone a Forbush decrease with an amplitude of about 4 per cent at the high latitude neutron monitor ground stations. At that time the observed integral flux of primary protons in energies greater than 350 Mev was slightly decreased above the values of August 22 and

September 15. It could be shown that, after the subtraction of flare particles, the flux of galactic protons with $E > 350$ Mev had decreased by at most 9 per cent. This is consistent with the station data.

In order to investigate the behavior of the primary electrons during the Forbush decrease we give in Figure 3 the altitude dependence of the electron events for the three balloon flights.

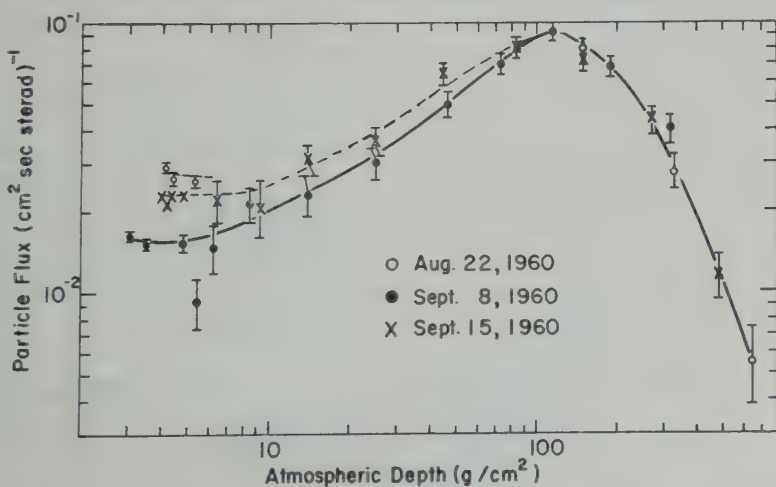


Fig. 3

Fig. 3. Vertical flux vs. atmospheric depth for minimum-ionizing particles with range between 10.5 and 122 g/cm² of lead (error limits shown are standard deviations).

The graph shows the flux of electrons with energies ranging from approximately 100 to 1000 Mev including the contribution of high-energy protons or mesons which fake an 'electron' event by making a nuclear interaction in the lead absorber.

It is found that, under 3 to 5 g/cm² of air, where the contribution of secondary electrons becomes insignificant, the flux decreased by 43 per cent between August 22 and September 8 and recovered to a level close to the August 22 value on September 15. Toward larger atmospheric depth the contribution of secondary electrons increases rapidly. Near and below the transition maximum, which is dominated by secondary electrons, no significant intensity changes can be observed. This is in agreement with the fact that the proton flux with energies above 350 Mev, which is the prime source of the secondary electrons, changed very little between the flights. It is, in addition, an independent proof that we are indeed observing primary electrons near the top of the atmosphere.

Figure 2 shows the flux of protons in the rigidity interval from 370 to 890 Mv and of electrons (uncorrected) in the rigidity interval from about 100 to about 1000 Mv. On September 8 the proton flux in the above rigidity interval is greatly enhanced due to storage of the solar flare particles from September 3. In a similar rigidity interval the electron flux is reduced as a consequence of the Forbush decrease. The uncorrected electron flux contains a background caused by interacting near relativistic protons. Since we know that the flux of these protons remained almost unchanged in the three measurements [Vogt, 1961] a subtraction of its contribution to the electron events would enhance the amplitude of the Forbush decrease observed in the electron component. Therefore, the 43 per cent change of the electron flux is a lower limit.

We draw the following conclusions from these observations:

1. If, during the September 3 flare, electrons

were produced on the sun, they were not stored in interplanetary space like protons of similar rigidity. Since this does not appear plausible the measurement indicates that the flare did not lead to an emission of electrons.

2. A Forbush decrease, which began on September 4, 1960, and amounted to about 4 per cent at high-latitude neutron monitor station and to about 9 per cent for primary protons with $E > 350$ Mev, affected the primary electron component and reduced the electron flux by more than 40 per cent in the rigidity range from 100 to 1000 Mv. This strongly suggests that the electrons observed at the earth have a history that is similar to the galactic protons. The evidence, therefore, points to a galactic origin of the primary electrons. However, it does not exclude the possibility of solar emission with subsequent, temporary storage in interplanetary space at other times.

Acknowledgments. We are deeply indebted to Mr. T. Burdick for his invaluable assistance throughout the experiment and to Mr. G. Lee who carried out the computer analysis. This research was supported in part by the National Science Foundation (grants nos. NSF-G7878 and NSF-G14889), the Office of Scientific Research, ARDC, United States Air Force (contract no. AF 18(600)-666) and by the Office of Naval Research, Skyhook Program (grant no. Nonr-(C) 00010-60).

REFERENCES

- Davis, L. R., C. E. Fichtel, D. E. Guss, and K. Ogilvie, Rocket observations of solar protons on September 3, 1960, *Phys. Rev. Letters*, **6**, 492, 1961.
- Earl, J. A., Cloud-chamber observations of primary cosmic ray electrons, *Phys. Rev. Letters*, **6**, 125, 1961.
- Meyer, Peter, and Rochus Vogt, Electrons in the primary cosmic radiation, *Phys. Rev. Letters*, **193**, 1961.
- Vogt, Rochus, Primary cosmic ray and solar protons, *Phys. Rev.*, 1962.
- Winckler, J. R., P. D. Bhavsar, A. J. Masley, and T. C. May, Delayed propagation of solar cosmic rays on September 3, 1960, *Phys. Rev. Letters*, **6**, 488, 1961.

(Received September 5, 1961.)

Geomagnetic Activity and the Reception of Whistlers in Polar Regions

G. McK. ALLCOCK AND M. F. RODGERS

*Dominion Physical Laboratory
Lower Hutt, New Zealand*

The correlation between the occurrence of whistlers and geomagnetic activity has often been sought. Usually the whistler data from middle-latitude stations have been used, and the results have been inconclusive or even conflicting. This may be attributed at least partly to the use of the 'whistler rate' (the number of whistlers heard per unit time) as a suitable characteristic for comparison with geomagnetic activity indices. The whistler rate is, however, likely to be strongly dependent on the prevailing meteorological conditions; a large whistler rate may occur because a rapid succession of lightning flashes occurs in a suitable geographical area. The presence, during a recording schedule, of only one whistler is sufficient to establish that whistler propagation path is open; a larger number of whistlers during the schedule merely indicates the existence of a larger number of suitable lightning flashes. Thus a more profitable approach to the problem appears to be that of considering the percentage of schedules (normally of 2 minutes' duration each hour) during which whistlers are heard, irrespective of their number or strength. This is less likely to be influenced by meteorological conditions, and more likely to be a measure of whistler propagation conditions. Using this approach, Barrington [1960] has found a negative correlation between whistler reception and geomagnetic activity for stations in the geomagnetic latitude range 50° to 66° , the effect tending to become more pronounced as the auroral zone is approached. It is the purpose of this paper to report that the whistler data from Scott Base, Ross Dependency, Antarctica (geomagnetic latitude 79°S), situated on the polar edge of the auroral zone, show an even more marked inverse relationship to geomagnetic activity. The effect of the diurnal variation of geomagnetic activity on the shape of the diurnal curve of whistler reception is also discussed. Between March and December 1959, record-

ings of whistlers were made at Scott Base on 2 days a week for 2 minutes in each hour, according to an international agreement. A preliminary analysis of the data showed that in all months there was a strong diurnal variation in whistler reception. Nevertheless, there did not appear to be any significant seasonal variation in the shape of the diurnal curves. Thus, to keep the sample sizes as large as possible, no division of the data into seasonal groups was made when a correlation with geomagnetic k indices was sought.

Since k indices are based on magnetic conditions during 3-hourly intervals, the whistler data were divided into 8 similar groups, according to the time of day. For example, the recording schedules commencing at 0035, 0135, and 0235 UT were grouped together for comparison with the k index for the period 00h to 03h UT on the same day. These were then classified according to the presence or absence of audible whistlers and tabulated against k index, as in Table 1. For this particular group the correlation coefficient was calculated to be -0.21 . Data for the other seven groups gave correlation coefficients varying between -0.16 and -0.30 , all of which are significant at the 1 per cent level.

TABLE 1. Scott Base Whistler Data (March–December 1959) for Schedules Commencing at 0035, 0135, and 0235 UT

k index (00–03 UT)	0	1	2	3	4	5	6	7
Number schedules containing audible whistlers	0	5	10	13	0	3	0	0
Number schedules lacking audible whistlers	0	7	23	79	25	11	0	3
Percentage schedules containing audible whistlers	–	42	30	14	0	27	–	0

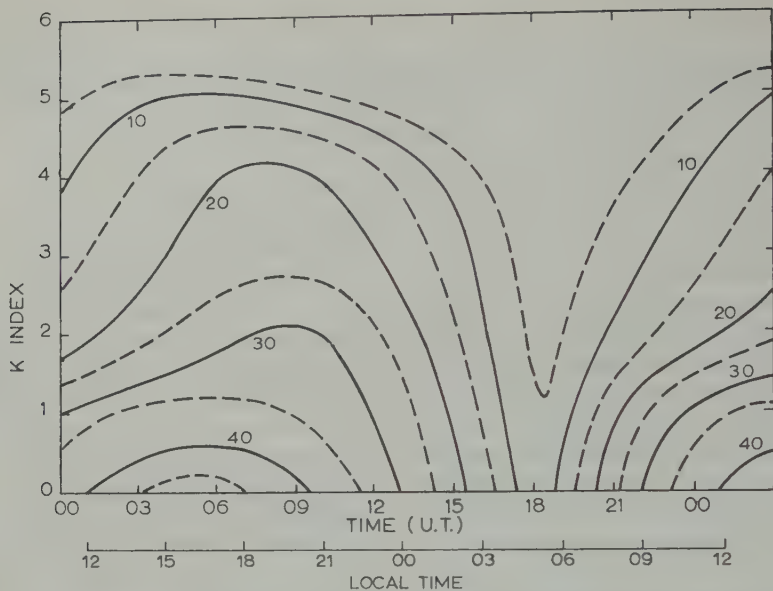


Fig. 1. Curves of equal whistler reception (percentage of schedules containing audible whistlers) at Scott Base during 1959, as a function of local k index and time of day.

The results from the eight tables are combined and plotted as Figure 1, which shows how the percentage of schedules containing audible whistlers varied with both k index and time of day. (Some smoothing of the original percentage values was introduced when the curves in Figure 1 were drawn; that such smoothing is insignificant is indicated by the fact that the correlation coefficient between the original values and those read off Figure 1 is $+0.86$.) It will be seen that at all times of day there was a marked decline in whistler reception as the k index increased. This decline was of such a magnitude that for average geomagnetic conditions ($k = 3 \pm 1$) the percentage of schedules containing audible whistlers was only about one-half that obtaining for completely quiet conditions ($k = 0$). On none of the 44 occasions during this 10-month period when the k index was 6 or greater was any whistler heard at all.

The percentages shown in Figure 1 are likely to be minimum values, for the following reasons: (a) on occasions an appreciable amount of man-made interference would have inhibited the detection of very weak whistlers; (b) for schedules during which no whistlers were detected, there is no certainty that suitably placed whistler-producing lightning discharges did in fact take place. To this extent, the above

method of analysis is still not completely free from the influence of meteorological conditions. This uncertainty can be eliminated if, instead of observing whistlers, we observe man-made VLF transmissions propagated in the whistler mode, as described by Helliwell and Gehrels [1958].

The above result extends and gives quantitative support to earlier observations by Marti [1958; 1961] that on most occasions of strong whistler activity at Scott Base local geomagnetic activity was moderate or quiet. A similar observation has been made by Ungstrup [1959] for whistler reception at Godhavn, Greenland (geomagnetic latitude 79.8° N).

Using Figure 1, the effect of the diurnal variation of geomagnetic activity can be taken out of the whistler data. For the raw data, i.e. for average geomagnetic conditions existing in 1959, the maximum whistler reception occurred near 08h UT with a minimum near 19h UT. However, Figure 1 shows that, for completely quiet geomagnetic conditions, the maximum would have occurred about 3 hours earlier, whereas the minimum would have occurred about 1 hour earlier. Thus the times of maximum and minimum whistler reception in polar regions will vary from year to year, in accordance with changes in average geomagnetic activity. It may

Therefore be necessary to take into account the observations when making comparisons of whistler reception at different stations in high latitudes.

Acknowledgment. This work forms part of the radio research program of the Dominion Physical Laboratory. We wish to acknowledge the benefit of helpful discussions with colleagues.

REFERENCES

Arrington, R. E., Contribution to discussion on 'IGY Whistler Results' at 13th General Assembly of URSI, London, 1960.

Helliwell, R. A., and E. Gehrels, Observations of magneto-ionic duct propagation using man-made signals of very low frequency, *Proc. IRE*, **46**, 785-787, 1958.

Martin, L. H., Whistlers in the Antarctic, *Nature*, **181**, 1796, 1958.

Martin, L. H., Whistlers, *Scott Base, New Zealand IGY Antarctic Expeditions, Scott Base and Hallett Station*, NZDSIR Bulletin 140, pp. 123-126, 1961.

Ungstrup, E., Observations of whistlers and very low frequency phenomena at Godhavn, Greenland, *Nature*, **184**, 806, 1959.

(Received May 5, 1961; revised August 7, 1961.)

Discussion of Paper by John R. Spreiter and Benjamin R. Briggs on ‘Theory of Electrostatic Fields in the Ionosphere at Polar and Middle Geomagnetic Latitudes’

D. T. FARLEY, JR.¹

Chalmers University of Technology, Gothenburg, Sweden

In their paper *Spreiter and Briggs* [1961] presented new computations of the electrostatic coupling between the *E* and *F* regions, based on ionospheric data much more extensive and reliable than that used in my earlier calculations [Farley, 1959; 1960]. The numerical results obtained were similar to mine but indicated an even stronger degree of coupling. In discussing the boundary condition at infinity to be used in the calculation, Spreiter and Briggs made some points that I believe could be debated, but since any differences of opinion that we may have are rather academic and do not affect the final answer, there seems to be little point in discussing them here. I certainly agree with their final results, which undoubtedly will be of interest in connection with various ionospheric problems.

I do differ, however, with one of the conclusions reached by Spreiter and Briggs, namely, the conclusion that Dagg's [1957] suggestion concerning the cause of the irregularities of electron density in the *F* region (the suggestion that the irregularities responsible for radio-star scintillation and spread *F* may be produced by irregular electrostatic fields resulting from turbulent winds in the dynamo region) may well have some merit. On the basis of the electrostatic coupling calculation alone, this is a reasonable conclusion and is the one also arrived at by Farley [1959]. However, as Spreiter and Briggs point out, one must also consider the mechanism by which the irregular electric fields are produced.

Some work on this question has already been described [Farley, 1960]. These calculations were certainly, for the most part, very approximate, but even a crude investigation shows that Dagg's suggestion is a dubious one at best. For instance, at the magnetic poles, as was also

pointed out by Dougherty [1959], no density fluctuations at all can be produced by an electric field whose curl is zero. The drift velocity of the ionization is simply $E \times B/B^2$, and the divergence of this quantity is identically zero if *B* is a constant and $\text{curl } E = 0$.

At lower latitudes the small-scale drift motions will have a vertical component, and one might expect that some irregularities could be produced because of the vertical gradient of electron density in the ionosphere. However, this effect was found to be a very weak one, primarily because any small-scale source field will be nearly short-circuited within the dynamo region itself. If the source is considered to be in the upper part of the dynamo region, so that the coupling to the *F* region is strong, the source itself will be weak because of the near short circuit existing below it. In general, this short-circuiting effect tends to increase as the coupling increases, and so the stronger coupling found by Spreiter and Briggs will probably not lead to any significant enhancement of the irregular vertical drifts.

To be sure, it would be desirable to examine this question in more detail, but it seems doubtful that the main conclusions would be changed and certainly the complete failure of Dagg's mechanism near the poles is a serious flaw. Thus it would appear to be necessary to search for some other mechanism to explain the formation of the *F*-region irregularities.

REFERENCES

- Dagg, M., The origin of the ionospheric irregularities responsible for radio-star scintillations and spread-*F*, 2, Turbulent motion in the dynamo region, *J. Atmospheric and Terrest. Phys.*, **1**, 139-150, 1957.
- Dougherty, J. P., Magnetohydrodynamics of the small-scale structure of the *F* region, *J. Geophysical Research*, **64**, 2215-2216, 1959.
- Farley, D. T., Jr., A theory of electrostatic field

¹ Now with the National Bureau of Standards, Boulder, Colo.

a horizontally stratified ionosphere subject to a constant, vertical magnetic field, *J. Geophys. Research*, *64*, 1225-1233, 1959.

W. D. T., Jr., A theory of electrostatic fields in the ionosphere at nonpolar geomagnetic latitudes, *J. Geophys. Research*, *65*, 869-877, 1960.

Spreiter, John R., and Benjamin R. Briggs, Theory of electrostatic fields in the ionosphere at polar and middle geomagnetic latitudes, *J. Geophys. Research*, *66*, 1731-1744, 1961.

(Received July 22, 1961; revised August 10, 1961.)

Increase of Ionization Associated with Geomagnetic Sudden Commencements¹

S. MATSUSHITA

*High Altitude Observatory, University of Colorado
Boulder, Colorado*

It has been suggested in previous studies on geomagnetic sudden commencements that charged particles from outside the earth's atmosphere seem to penetrate into the lower ionosphere at high latitudes and cause ionization there at the moment of sudden commencements [Matsushita, 1957; 1960]. Riometer records of 27.6 Mc/s cosmic noise obtained at Fort Yukon (geographic 66°34'N, 145°18'W), Barrow (geographic 71°31'N, 156°20'W), and Thule (geographic 76°33'N, 68°50'W), which are published in the IGY Aurora (Instrumental) Report, No. 1, Geophysical Institute, University of Alaska, clearly support the above suggestion.

As is seen in Figure 1, an absorption of 0.1 to 5 db often begins at the same time as the sudden

commencement and lasts 5 to 10 minutes. The time and date of all obvious sudden commencements in 1958 (all those reported by more than 30 geomagnetic stations in the world) and the occurrence of the absorption phenomenon are listed in Table 1. The asterisk in column 'PCA' indicates that polar-cap absorption due to solar flares preceded the sudden commencement; and the PCA in column 'Increase of Absorption' means that polar-cap absorption is still continuing at the time of the sudden commencement. It can be noticed in Table 1 that the increases of absorption occur more frequently at Barrow and Fort Yukon than at Thule; in other words, more in the auroral zone than in the polar cap. No absorption increases were found at the time of the less obvious sudden commencements, which are reported by fewer than 30 geomagnetic stations in the world.

The shape of sudden commencements in the

¹ Presented at the Pacific Southwest Regional Meeting of the American Geophysical Union, Berkeley, California, January 1961.

TABLE 1. All Obvious Sudden Commencements in 1958 and Accompanying Increases of Absorption in Riometer Records

Sudden Commencements		Type	PCA	Increase of Absorption		
1958	UT			Thule	Barrow	Fort Yukon
Feb. 11	01:25	SC	*			PCA
	16 16:42	-SC				x
Mar. 3	09:31	SC				0
	14 12:12	SC				x
	25 15:40	SC	*			PCA
Apr. 26	12:47	SC			0	0
May 31	16:52	SC		0	X	
June 7	00:46	SC	*	0	0	
	14 18:28	SC		0	0	
	28 07:13	SC		0	0	
	28 17:42	SC		0	0	
July 8	07:48	SC	*	PCA	PCA	
	21 16:37	SC			X	
	31 15:29	SC	*	0	0	
Aug. 17	06:22	-SC	*	PCA	X	
Sudden Commencements		Type	PCA	Increase of Absorption		
1958	UT			Thule	Barrow	Fort Yukon
Aug. 22	02:27	-SC				
	24 01:40	SC	*	0	X	
Sept. 3	08:43	SC		0	0	0
	16 09:30	SC		0	0	0
	25 04:08	-SC	*	0	PCA	PCA
	30 10:05	SC		-	x	x
Oct. 22	03:15	-SC		0	X	
	24 07:30	SC		0	X	
	27 15:23	SC		-	0	
	28 06:50	SC		0	0	
Nov. 28	01:09	-SC		0	-	
Dec. 4	00:35	-SC		0	x	
	13 00:01	-SC		0	x	
	15 20:22	-SC		X	x	
	17 15:47	SC		-	0	

0, No occurrence.

-, Interference.

X, Certainly occurred.

x, Probably occurred.

BARROW

Geographic 71.5° N. 156.3° W.; Geomagnetic 68.5° N. 241.3° E.

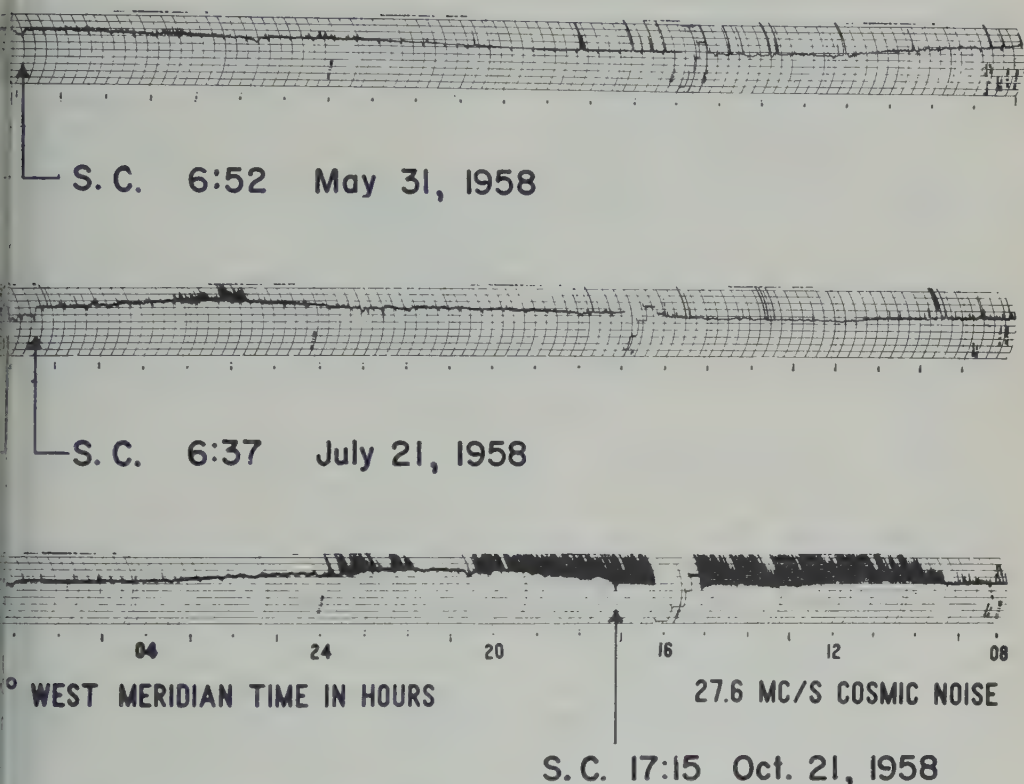


Fig. 1. Three examples of increase of absorption, which begin at the same time as the sudden commencement, in riometer records of 27.6 Mc/s cosmic noise obtained at Barrow. The time scale is from right to left.

horizontal geomagnetic variation field can be classified into three types [Matsushita, 1960]; SC is characterized by a small negative impulse preceding the main positive impulse, SC- is a common main positive impulse alone, and SC+ is characterized by an increase lasting from 0 to about 6 minutes followed by a decrease to a level lower than the initial pre-sudden commencement level. A combination of SC- and SC+, which is a negative impulse preceding the main positive impulse followed by a sharp increase, occasionally occurs; it is indicated by SC-. These types, based on geomagnetic data obtained at College and Barrow, Alaska, are listed in Table 1. SC- (including SC-), SC-, and SC+ are accompanied by six, five, and one

cases of increased absorption, respectively, at Barrow and Fort Yukon. When we eliminate the events in which polar-cap absorption is still continuing and interference occurs, we have 100, 45, and 14 per cent occurrence frequency of increased absorption accompanying the types SC- (including SC-), SC-, and SC+, respectively. Thus, SC- almost always seems to be accompanied by increases of cosmic noise absorption.

Ionospheric variations as observed by vertical sounding ionosondes soon after the time of the increased absorption associated with sudden commencements are listed in Table 2. Beginning of blackout and blanketing of the F region by E_s echoes, increase of f_oE_s and f_{min} , and appearance of the E_s often occurred; all these

TABLE 2. Ionospheric Variations at the Time of Increases of Cosmic Noise Absorption Associated with Sudden Commencements

Sudden Commencement	1958	UT	Increase of Absorption at Thule	Ionospheric Variation at Thule	Increase of Absorption at Barrow	Ionospheric Variation at Barrow	Increase of Absorption at Ft. Yukon	Ionospheric Variation at College
Feb. 16	16:42						x	NO
Mar. 14	12:12						x	Blanketing of F begins, and f_{min} increases
May 31	16:52				X	Blanketing of F begins, and E_s appears		
July 21	16:37				X	Blanketing of F begins, and E_s increases		
Aug. 17	06:22				X	B		
	24 01:40				X	C		
Sept. 30	10:05				x	Blackout begins	x	NO
Oct. 22	03:15				X	f_{min} slightly increases		
	24 07:30				X	B		
Dec. 4	00:35				x	f_{min} increases		
	13 00:01				x	f_{min} increases		
	15 20:22	X	E_s appears		x	f_{min} slightly increases		

X, Certainly occurred.
x, Probably occurred.

B, During blackout.
C, Failure of ionosonde.

phenomena indicate an increase of ionization in the E layer and/or an increase of radio absorption in the D region [Matsushita, 1958]. In other words, charged particles (probably high-energy electrons) and/or X rays produced by bremsstrahlung from the high-energy electrons seem to penetrate into the lower ionosphere from outside the earth's atmosphere and to cause ionization, and, as a consequence, increased absorption of radio wave and cosmic noise at the time of sudden commencement. In fact, a burst of X rays was observed at balloon altitude over Alaska, simultaneous with increases of ionospheric absorption in Alaska, Sweden, and Norway, at the onset of the sudden commencement of June 27, 1960 [Brown, Hartz, Landmark, Leinbach, and Ortner, 1961].

The amount of increased ionization at these times can be estimated by making use of the relation

$$\int k \, dS = 0.461 \int N\nu / \{ \nu^2 + (\omega \pm \omega_L)^2 \} \, dS \tag{1}$$

where N is electron number density per cubic centimeter, ν is effective electron collision fre-

quency per second (10^6 at about 100 km height and 10^6 at about 85 km), ω_L is gyro-magnetic frequency of electron due to the longitudinal component of the earth's magnetic field, and ω is the observing frequency, i.e., $\omega = 2\pi(27.6 \times 10^6) > \omega_L \geq \nu$. The sudden commencement absorption effect in decibels, $\int k dS$, is assumed to be 1 db on the average. The integration in equation 1 should be carried out over the effective absorbing region. If the thickness of the region is taken as 10 km, N is 6.5×10^6 per cm³ at about 100 km altitude and 6.5×10^4 per cm³ at about 85 km, assuming ν does not change in the lower ionosphere these N become roughly ten times larger than normal.

Owing to a coupling effect of the earth's magnetic field and the solar plasma, which is responsible for the geomagnetic storm, hydromagnetic waves will be formed. As suggested previously [Matsushita, 1960], these waves, which propagate along the earth's magnetic lines of force, may cause charged particles in the outer Van Allen belt to move down along the lines of force. Malville [1961] calculated acceleration of Van Allen electrons by electrostatic waves propagating parallel to the magnetic lines of force and showed a possibility of bursts

electrons caused by a decrease of pitch angle of trapped electrons due to those electrostatic waves. In any case, Van Allen electrons seem to play an important role in the present absorption phenomenon, because an increase of ionization begins at the time of sudden commencement, and at the time of preceding negative impulses of sudden commencements. Details will be discussed in a full paper with more observational results using the IGY Aurora (Instrumental) report, No. 2.

Acknowledgments. I wish to express my sincere gratitude to Dr. W. O. Roberts, Mr. A. H. Shapley, and R. W. Knecht, and Dr. H. Leinbach for their kind help, and I also want to thank the Boulder Laboratories of the National Bureau of Standards for extending their facilities to me under an appointment as a guest worker. The work reported here was supported by the National Academy of Sciences under grant NSF-G9427.

REFERENCES

- Brown, R. R., T. R. Hartz, B. Landmark, H. Leinbach, and J. Ortner, Large-scale electron bombardment of the atmosphere at the sudden commencement of a geomagnetic storm, *J. Geophys. Research*, **66**, 1035-1041, 1961.
- Malville, J. M., The interaction of plasma streams with the terrestrial exosphere (chapter 5 of Ph.D. thesis, Studies of fast-drift radio bursts and related phenomena), Astro-Geophysics Department, University of Colorado, 1961.
- Matsushita, S., On sudden commencements of magnetic storms at higher latitudes, *J. Geophys. Research*, **62**, 162-166, 1957.
- Matsushita, S., Some studies of the upper atmosphere in the auroral zone, *Ann. géophys.*, **14**, 483-491, 1958.
- Matsushita, S., Studies on sudden commencements of geomagnetic storms using IGY data from United States Stations, *J. Geophys. Research*, **65**, 1423-1435, 1960.

(Received June 5, 1961; revised September 12, 1961.)

Two Methods of Detecting Ionospheric Disturbances

G. A. M. KING AND M. D. LAWDEN

Geophysical Observatory, Christchurch, New Zealand

Methods of detecting ionospheric disturbances using analyses of routine numerical data are in general use. They have shortcomings, however, and it was considered desirable to test methods based on a visual representation of ionospheric records covering a large period of time relative to the durations of the disturbances. These can give a better idea of the character of a disturbance, and there is room for controlled use of the observer's judgment. As no universal agreement has been reached on what constitutes a disturbance, for the purpose of this investigation a disturbance was taken to mean a sudden change in F -region conditions, often of small degree and lasting a few hours.

In mid-latitudes during summer months, ionospheric disturbances are characterized by an unusually high $h'(f)$ trace and an unusually low f_oF_2 . These characteristics form the basis of two methods for storm detection, which may be

called the displacement (D) method and the virtual-height (V. H.) method. In the method, suggested by King's [1960] overlapping trace technique, the ordinary traces of the F_2 region were copied onto grids of frequency and height lines plotted from one of the ionogram frames. A grid was assigned to each hour, and on it were copied all the traces that occurred at that hour over the period of a month. See Figure 1.

Assuming that calm conditions normally prevail, the disturbed records will stand out conspicuously above the rest. After the grids had been filled with a month's traces the film was gone through again, each frame being projected on its corresponding grid in sequence and significant changes in the position of the trace investigated. A sharp change in relative position indicated the onset of a disturbance.

In the V. H. method, $h'F$ was plotted against time for several days and disturbed periods were indicated by the graph rising above the normal level. See Figure 2. Due care was taken to allow for sporadic E .

Since the D method is more effective for daylight hours when there is much variation in the visual characteristics of the F_1 - F_2 transition region, in this investigation the D method was used during the daytime (0600-1800) and the V. H. method at night. The records analyzed were from Christchurch data for November 1951.

As a result of the survey 18 disturbances were found. This large number indicates that most of them were of small amplitude and probably represent the limit of observation beneath which they have no significance.

Because there is believed to be a correspondence between ionospheric and magnetic disturbances, the magnetic records for Christchurch were examined for the same month. The times of occurrence of baylike disturbances in H were noted. When the occurrences of ionospheric

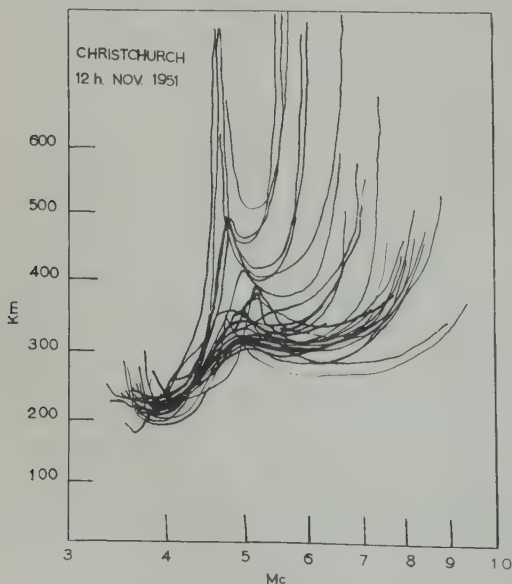


Fig. 1. The D method—a typical completed grid. Christchurch, November 12, 1951.

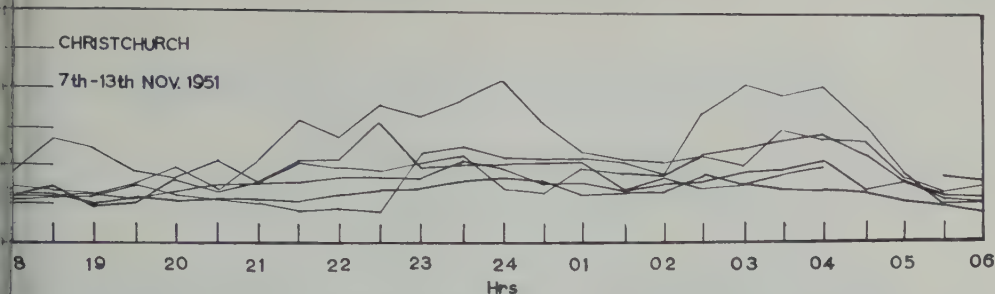


Fig. 2. The V. H. method—a typical completed grid. Christchurch, November 7–13, 1951.

magnetic disturbances were compared, 14 times were common to both (within table limits). The magnetic records suggest that some ionospheric disturbances might have occurred that were not at first detected. As a result of further examination 9 more small ionospheric disturbances were found to occur during periods of magnetic activity, bringing the number of correlated occurrences up to 23. These results may be tabulated:

Correlated Ionospheric and Magnetic Disturbances	
Using the D method (day)	10
Using the V. H. method (night)	13
Total	23

Uncorrelated Ionospheric and Magnetic Disturbances

Ionospheric disturbances	4
Magnetic disturbances	5

Both the D method by day and the V. H. method by night proved effective for detecting small ionospheric disturbances in middle latitudes during the summer.

REFERENCE

King, G. A. M., A convenient method of getting representative ionospheric heights, *J. Geophys. Research*, 65, 1623–1624, 1960.

(Received June 5, 1961; revised August 11, 1961.)

Note on 'Vertically Traveling Shock Waves in the Ionosphere'

FRED B. DANIELS AND ARTHUR K. HARRIS

*U. S. Army Signal Research and Development Laboratory
Fort Monmouth, New Jersey*

In an earlier communication [Daniels, Bauer, and Harris, 1960] we reported on vertically traveling shock waves in the ionosphere that were caused by atomic explosions at the earth's surface. These shock waves distorted the h' - f trace of an ionospheric recorder in a manner similar to that resulting from the down-coming disturbances described by Bibl and Rawer [1959], with the exception that the latter had a much smaller vertical velocity component. This discrepancy led us to re-examine our results, and we found that on one occasion the disturbance consisted of three arrivals, one of which had a much lower velocity than the other two. (Unfortunately the original records of the other tests that evidenced multiple arrivals were not available for re-examination.) The exact velocity of the late arrival in this one instance could not be determined because of radio-frequency interference that blanketed a large part of the h' - f trace. The velocity was, however, of the order of magnitude of that reported by Bibl and Rawer (115 m/sec). In view of this new finding, our earlier identification of the second arrival as the 'retarded sound wave' is probably incorrect, and this arrival should be considered to be an ordinary hydrodynamic shock wave. In fact, any classification of the arrivals analogous to the three hydro-magnetic components would be very difficult because of the variety of modes that may be present if all kinds of particles and nonlinear as well as linear oscillations are considered.

An approximate value for the pressure amplitude of the upgoing hydrodynamic shock wave

(the second arrival) can be computed from velocity. This amplitude could be adopted as a very rough estimate of that of the downcoming slow wave, since both waves appear to be caused about the same amount of distortion in the h' - f trace. Such an estimate indicates that the amplitude of the disturbances observed by Bibl and Rawer may be large enough to be detectable at the earth's surface by acoustic methods. In fact, the downcoming disturbances observed by Bibl and Rawer may be identical to the traveling pressure waves that have been observed with infrasonic microphones by Chazanowski, Greene, Lemmon, and Young [1960]. We therefore suggest that, in future experimentation along these lines, ionosphere recorders be operated continuously while acoustic observations are being made, to see whether individual acoustic and ionospheric arrivals are correlated.

REFERENCES

- Bibl, K., and K. Rawer, Traveling disturbances originating in the outer ionosphere, *J. Geophys. Research*, **64**, 2232-2238, 1959.
 Chazanowski, P., G. Greene, K. T. Lemmon, J. M. Young, Traveling pressure waves associated with geomagnetic activity, Presented at meeting of American Geophysical Union, December 27-30, 1960. Abstract in *J. Geophys. Research*, **65**, 2482, 1960.
 Daniels, F. B., S. J. Bauer, and A. K. Harris, Vertically traveling shock waves in the ionosphere, *J. Geophys. Research*, **65**, 1848-1860, 1960.

(Received August 12, 1961.)

Discussion of Incoherent Backscatter Power Measurements at 440 Mc/s

V. C. PINEO AND H. W. BRISCOE

*Lincoln Laboratory,¹ Massachusetts Institute of Technology
Lexington 73, Massachusetts*

Millstone Hill backscatter power measurements reported by *Pineo, Kraft, and Briscoe* can be misinterpreted to indicate an agreement with theoretical estimates that are based on a scattering coefficient equal to the square of the classical electron radius. The observed backscattered power was, however, much less than these theoretical estimates.

To clarify the discussion of these measurements, the equation that was used with the experimental data to calculate the scattering cross section is given below.

$$P_r = \left(\frac{P_t A}{4\pi R^2} \right) \left(\frac{C\tau}{2} \right) \left(\frac{1}{L} \right) (N\bar{\sigma}) \quad (1)$$

Equation 1 is the familiar radar equation for volume-scattered scatterers [*Kerr*, 1951].

In the Millstone Hill backscatter experiment the received power is computed using equation 1 and measured returns from a height of about 300 km and previously these were from the *F*-region peak.

The electron density N_e at the *F*-region peak is deduced from measurements of the *F*-region frequency, f_o , using the relation

$$N_e = \frac{\pi m f_o^2}{e^2} = 1.24 \times 10^4 f_o^2 \quad (2)$$

Equation 2 can be found in any good textbook on the ionosphere (for example, *Mitra*, [1952]). It is based on the assumption that

$$N\bar{\sigma} = N_e \bar{\sigma} \quad (3)$$

where the value for $\bar{\sigma}$ was found to be about 7×10^{-30} (meter)².

It was pointed out that this value is approximately equal to the square of the classical electron radius. This observation, though true, is unfortunately misleading. As used in equation

1, the received power was computed with support from the U. S. Army, Navy, and Air Force.

1, $\bar{\sigma}$ is the radar cross section of the scatterers and should not be confused with the scattering coefficient σ_e of an electron which is defined by *Gordon* [1958] as the power scattered into a unit solid angle per unit incident power density per electron. Thus,

$$\bar{\sigma} = 4\pi\sigma_e \quad (4)$$

Gordon gave

$$\sigma_e = \left[\frac{\mu e^2}{4\pi m} \sin \psi \right]^2$$

where the units are mks. This has the same value as $(r_e)^2$. This result is also derived from electromagnetic theory by *Kerr* [1951].

The results of the Millstone backscatter experiment indicated a value for

$$\begin{aligned} \sigma_e &\simeq \frac{1}{4\pi} (7 \times 10^{-30}) (\text{meter})^2 \\ &\simeq 6 \times 10^{-31} (\text{meter})^2 \end{aligned}$$

These calculations of $\bar{\sigma}$ and σ_e from the experimental data did not take into account certain approximations involved in the derivation of equation 1 which for Millstone-type antennas cause the received power predicted for a given value of N to be about twice the actual received power. Taking these approximations into account results in apparent values for $\bar{\sigma}$ and σ_e which are twice the values given above, that is $\bar{\sigma} = 1.4 \times 10^{-29}$ (meter)² and $\sigma_e = 1.2 \times 10^{-30}$ (meter)².

These results have been confirmed by recent experiments in which bistatic measurements of forward-scattered signal strengths were compared with backscattered signal strengths from the same area in the *F* region (*Briscoe and Pineo*, 1961).

In the preceding discussion the observed scattered power was compared with *Gordon's* [1958] original prediction which was based on independent scattering by free electrons. More

recently developed theories by *Fejer* [1960a, b] (also private communication); by *Dougherty and Farley* [1960] (also private communication); and by *Hagfors* [1960] show that at radio wavelengths greater than the Debye length the electrons do not behave independently of each other, and that the scattered power is approximately one-half what it would be for scattering by free electrons.

In brief, then, the scattered power measurements at 440 Mc/s are about 7 db less than Gordon's predictions and about 4 db less than predictions based on the more recent theories of electron scattering.

REFERENCES

Briscoe, H. W., and V. C. Pineo, Forward propagation at 440 megacycles by incoherent scatter-

- ing in the *F* region, Paper read at May meeting of URSI.
 Dougherty, J. P., and D. T. Farley, *Proc. Soc.*, 259, 79-99, 1960.
 Fejer, J. A., Radio-wave scattering by an ionized gas in thermal equilibrium, *J. Geophys. Research*, 65, 2635-2636, 1960a.
 Fejer, J. A., *Can. J. Phys.*, 38, 1114-1133, 1960.
 Gordon, W. E., *Proc. IRE*, 46, 1824-1829, 1958.
 Hagfors, T., Scientific Report No. 1, Stanford Electronics Laboratory, 1960.
 Kerr, D. E., *Propagation of Short Radio Waves*, chap. 7, McGraw-Hill Book Company, New York, 1951.
 Mitra, S. K., *The Upper Atmosphere*, chap. 1, The Asiatic Society, Calcutta, 1952.
 Pineo, V. C., L. G. Kraft, H. W. Briscoe, S. characteristics of ionospheric backscatter observed at 440 Mc/s, *J. Geophys. Research*, 65, 2629-2633, 1960.

(Received June 14, 1961.)

Artificial Initiation of Lightning Discharges

M. BROOK

*New Mexico Institute of Mining & Technology
Socorro, New Mexico*

G. ARMSTRONG

*Museum of Science
Boston, Massachusetts*

R. P. H. WINDER

Commander USNR

B. VONNEGUT AND C. B. MOORE

*Arthur D. Little, Inc.
Cambridge, Massachusetts*

Investigations of atmospheric electricity could be considerably aided if one could cause lightning to strike from a storm at will. Were this possible, apparatus could be set up to observe and measure the discharge and determine the effects it produces.

It might be supposed that if a balloon or kite were flown on a wire near a storm, the probability of a lightning discharge to the wire would be high. Surprisingly enough, we have found that even when we have repeatedly flown balloons on several kilometers of piano wire directly into the storms over the summit of Mt. Washington, New Mexico, no stroke has ever struck the wire, even though discharges were occurring overhead at a rate of one or more per minute.

Our measurements show that currents of several milliamperes flow from the wire during a storm. We believe that the large amount of electric charge thus released reduces the electric field near the wire and screens it from discharges. Through this mechanism lightning rods probably prevent discharges under conditions of low wind velocity.

It has occurred to us and to others [Newman, 1957] that it might be possible to initiate a lightning stroke at a particular place and time by rapidly introducing a wire into the electric field of a storm by the use of a rocket or some other technique.

The hope of being able to trigger the discharge in this way, before a shielding region of space charge has had time to form, is strengthened by several chance observations of lightning strokes that appear to have been initiated acci-



Fig. 1. Lightning stroke apparently initiated by plume of water arising from a depth charge. (Courtesy of U. S. Naval Ordnance Laboratory.)

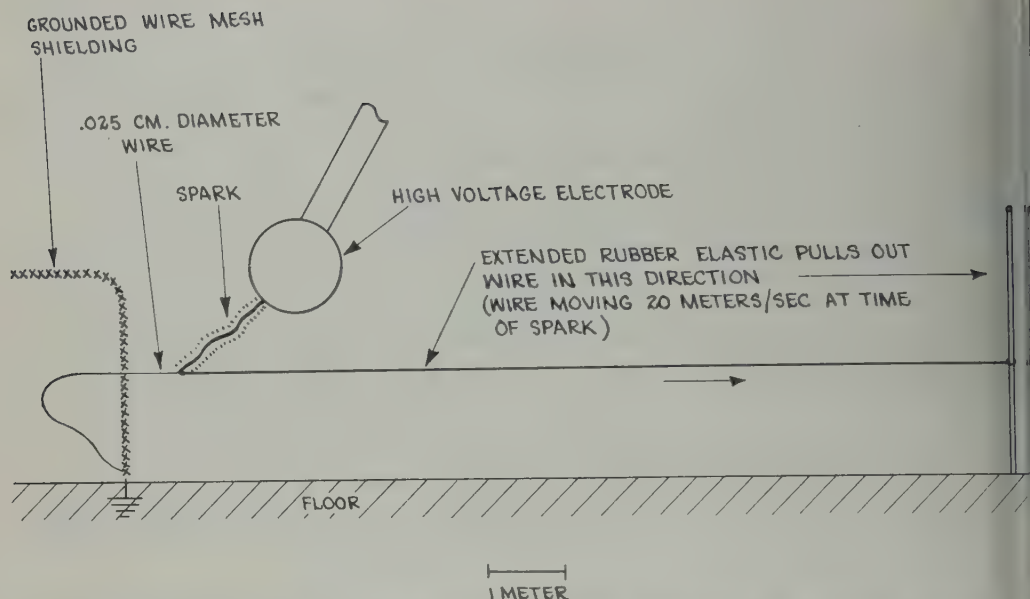


Fig. 2. Arrangement to pull wire suddenly into region of strong field beneath electrode.

dentally by man. Figure 1 shows a lightning stroke that was apparently triggered when a depth charge being tested by the U. S. Navy in Chesapeake Bay sent up a plume of water beneath a thunderstorm [Young, 1961]. The first element of the four-element discharge began when the plume of water had risen to about 70 m in about 1 second. We have learned that similar instances have been observed in which lightning has apparently been triggered by the water plume sent up by explosives set off for seismic investigations in the Gulf of Mexico (J. Connors, 1961, private communication).

We have recently had the opportunity of testing the practicability of triggering lightning with model experiments made possible through the cooperation of the personnel of the Museum of Science who let us make use of their Van de Graaff generator, one of the largest of its kind. The apparatus is arranged so that when it is turned on sparks about 3 meters long jump from the 1.2-meter-diameter electrode to ground.

We stretched horizontally a grounded stainless steel wire, .025 cm in diameter, under the electrode and about a meter above the floor and found that the apparatus no longer produced sparks. Instead, a corona discharge current flowed between the wire and the electrode. The electrical process acting on the wire was such

that it set the wire into a rotary motion similar to that of a skipping rope. Apparently, as the wire moved closer to the electrode, the point discharge current increased, thereby decreasing the field and allowing the wire to move away thus giving rise to a periodic motion.

Our second experiment was designed to determine whether we could initiate a discharge by introducing the wire rapidly into the electric field. We stretched a nonconducting elastic band across the building directly under the electrode. A length of grounded wire was fastened to one end of the stretched elastic so that when it was released the wire would be rapidly drawn through the screen shielding and out under the electrode. When we used this arrangement, as illustrated in Figure 2, we found that the sudden introduction of the wire invariably resulted in a spark such as is shown in Figure 3. From motion pictures of the operation we estimate that the wire was drawn into the region of high field at a rate of about 20 m sec^{-1} . It appears that the wire was moving rapidly enough so that the ionization produced by point discharge were not sufficient to reduce the field at the tip of the wire to below avalanche values.

On the basis of these simple experiments we conclude that the chances are good for triggering a lightning stroke by projecting a fine wire

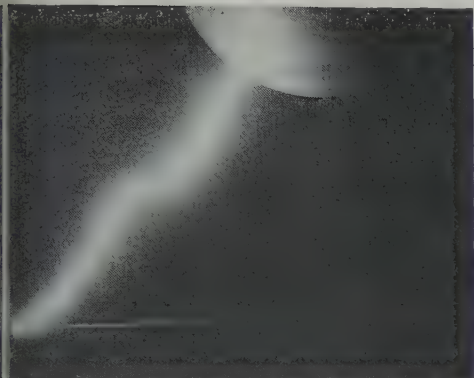


Fig. 3. High voltage spark jumping between vertical metal electrode and the fine wire that suddenly introduced. The wire entering from left is invisible; the white horizontal line to right of the lower end of the spark is the wire that is pulling the wire.

under a thunderstorm. This might be done with a bow and arrow, a kite or balloon with plastic, or by a rocket.

Acknowledgments. We are grateful to the staff

of the Boston Museum of Science for having let us make use of their Van de Graaff apparatus. We are pleased that the management has indicated its willingness to continue to make its unique and valuable high voltage apparatus available to all students of high voltage phenomena.

We thank the Commanding Officer of the Boston Branch, Office of Naval Research, for allowing Commander Winder to work on this project during his two weeks' annual training duty.

We wish to thank Mr. George A. Young of the U. S. Naval Ordnance Laboratory for supplying us data and photographs concerning the lightning striking the plume of water from the depth charge. We are pleased to acknowledge the support and aid of the Office of Naval Research and the National Science Foundation.

REFERENCES

- Newman, M. M., Lightning discharge channel characteristics, in *Recent Advances in Atmospheric Electricity*, edited by L. G. Smith, Pergamon Press, New York, p. 483, 1958.
- Young, G. A., A lightning strike of an underwater explosion plume, *U. S. Naval Ordnance Lab. Rept. 61-43*, to be published, 1961.

(Received July 17, 1961.)

Interplanetary Dust Particles of Micron-Size Probably Associated with the Leonid Meteor Stream

W. M. ALEXANDER, C. W. McCRACKEN, AND H. E. LaGOW

*Goddard Space Flight Center
National Aeronautics and Space Administration
Greenbelt, Maryland*

The IGY artificial earth satellite, 1959 Eta (Vanguard III), carried several sensors for determining the environmental conditions of the satellite. Among these sensors was a microphone for detecting individual impacts of interplanetary dust particles onto selected portions of the metallic shell of the satellite. In addition to contributing to the knowledge gained from direct measurements of interplanetary dust particles with rockets and other satellites about the model average mass and space distributions of small interplanetary dust particles, the dust-particle sensor on 1959 Eta has revealed an interval of unusual interplanetary dust particle activity. This letter is a report of the important details of the observed interplanetary dust-particle event, together with some tentative conclusions reached while analyzing the data.

The experiment data consist of a set of sequential counter readings that can be used along with a satellite ephemeris in order to determine impact rates for the interplanetary dust particles. Telemetry coverage of good signal-to-noise ratio was not sufficiently complete to allow a detailed segmentation of the satellite orbit into small sampling intervals. A good reading of the counter settings was obtained only about once per orbit, on the average.

The completed read-out of the telemetered data for the interval November 10-20, 1959, has shown the existence of a most interesting interplanetary dust particle event. This event was not evident in an earlier and preliminary data read-out, the results of which were reported by LaGow and Alexander [1960]. Impact rates averaged between available sequential static

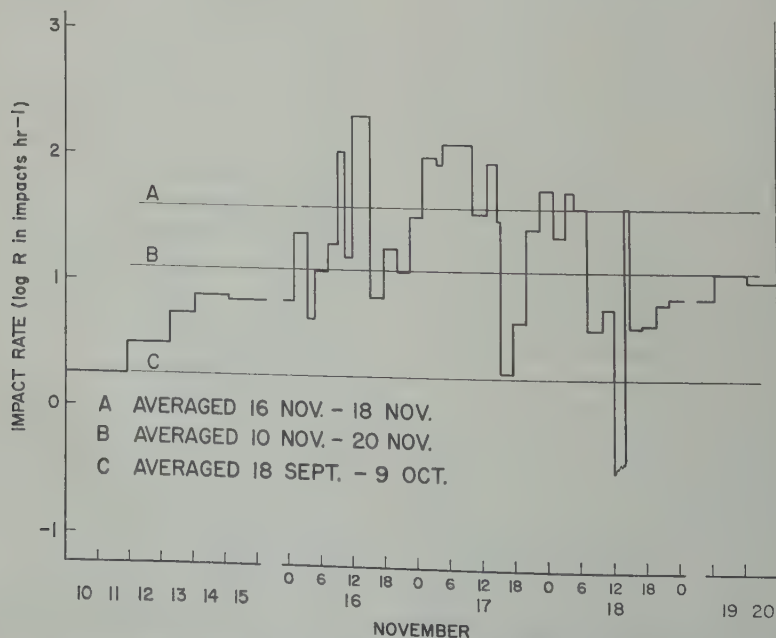


Fig. 1. Impact rates during the November 1959 interplanetary dust-particle event.

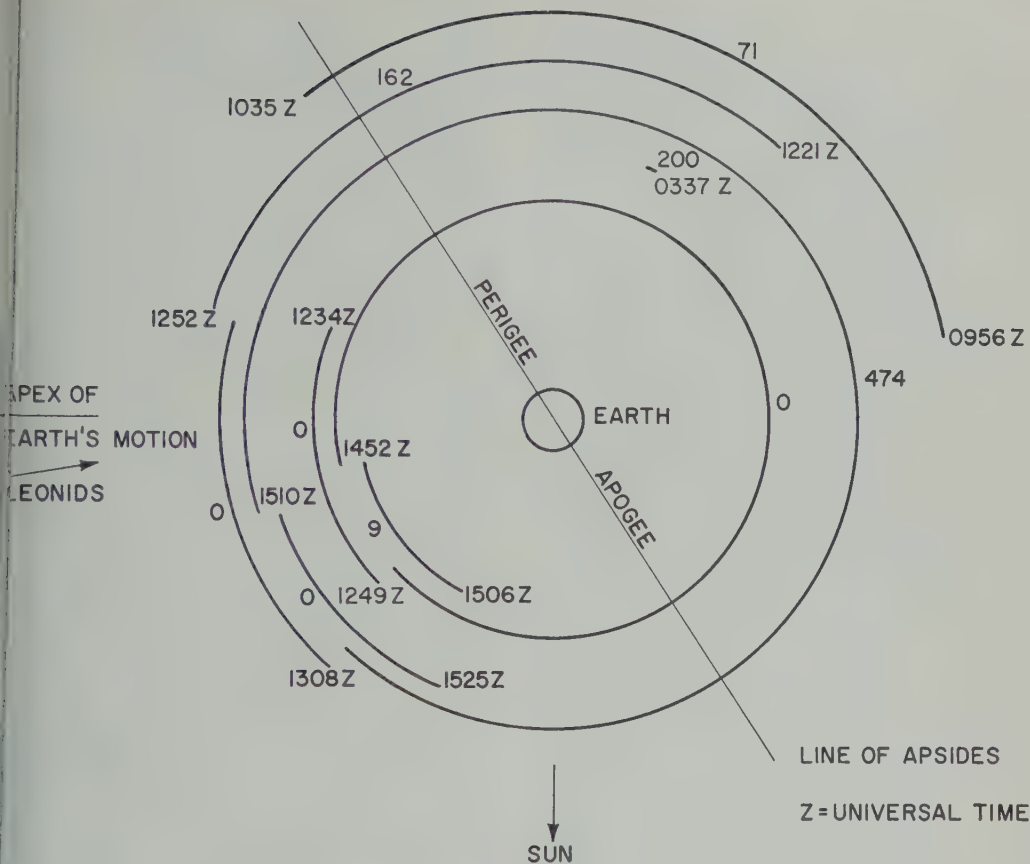


Fig. 2. Intervals of extreme activity during the November 1959 dust-particle event.

ses for a 70-hour interval on November 16-18 shown in Figure 1, together with average ly impact rates for the remainder of the interval November 10-20 on a compressed time le. The impacts of approximately 2800 dust particles with momenta greater than 1×10^{-2} ne sec were recorded during the 70-hour interval, which coincides closely with the time occurrence of the Leonid meteor shower. is number of impacts is approximately equal the number of impacts observed during the remainder of the lifetime (approximately 78 rs) of the experiment. The observed dust-ticle event certainly represents a pronounced, only short-lived, increase in interplanetary t-particle activity above the average back-und activity. Figure 1 serves also to show the idity with which the impact rate varied ing the event.

A representation of the orbital segments ing which extreme dust-particle activity

occurred on November 16, 17, and 18 is shown in Figure 2. The two segments of orbits during which unusually high impact rates (3 and 8 times the 70-hour average) occurred represent high velocity perigee passes when the satellite was moving almost directly into the Leonid radiant. In such cases, dust particles in the Leonid stream could have velocities relative to the satellite as high as 81 km sec^{-1} . Three orbital segments of approximately 15 minutes' duration each, during which no impacts occurred, all lay in mid-morning satellite local time, when the satellite was moving almost perpendicular to the direction of the Leonid radiant. At such times, dust particles in the Leonid stream would have had velocities relative to the satellite of about 70 km sec^{-1} . One of these intervals was the beginning of a 2.3-hour interval during which no impacts occurred. The 2.3-hour interval was followed immediately by a 14-minute mid-morning interval during which 9 impacts occurred.

In contrast to the average impact rate of 41.2 impacts per hour for the 70-hour interval are four sampling intervals of 1.3, 2.5, 2.5, and 2.3 hours' duration, during which the numbers of impacts were 7, 5, 13, and 0, respectively. These rates can be expressed in impacts per orbit as 12, 4, 11, and 0, respectively, which are considerably lower than the average of 39 impacts per orbit for the 70-hour interval.

The highest impact rate was 1900 impacts per hour when 200 ± 10 impacts were observed in 6.3 minutes of real-time telemetry at 2237 hours (satellite local time) on November 17. Although dust particles in the Leonid meteor stream would have had to pass tangentially through a portion of earth's atmosphere in order to have reached the satellite, the altitude and position of the satellite were such that the particles would have been negligibly decelerated by atmospheric drag before encountering the satellite. The 200 ± 10 dust particles could easily have come from the Leonid meteor stream.

The outstanding features of the interplanetary dust particle event, as described in the foregoing paragraphs are: (1) impact rates much higher than during the remainder of the lifetime of the experiment; (2) rapid fluctuations in the impact rate; (3) several orbits showing no or very few impacts; and (4) very high impact rates during several small segments of orbits and during an interval of real-time telemetry.

The dust-particle sensor was almost omnidirectional, making it difficult to determine much about the radiant of the dust particles. The shielding effect of earth could be used to an advantage in roughly determining the regions of the celestial sphere in which the radiant lay if better segmentation of the satellite orbit were possible. Correlation in time with the Leonid meteor shower of 1959, as observed by other means, is about the only means of determining whether the dust-particle event is possibly direct evidence of the presence of small dust particles in the Leonid meteor stream.

If the assumption is made that the majority of the interplanetary dust particles detected during the 70-hour interval were from the Leonid meteor stream, then an average particle velocity relative to the satellite may be assigned. The limits on particle velocity under this condition are 66 km sec⁻¹ and 81 km sec⁻¹. The value, 70 km sec⁻¹, is chosen as an appropriate

value to be applied during the 70-hour interval. The corresponding limiting mass sensitivity of the system is 1.4×10^{-9} g which for a mass density of 1 g cm⁻³ corresponds to a particle of 7 μ radius. The value 1 g cm⁻³ for mass density is somewhat higher than the value 0.05 g cm⁻³ used by Whipple [1958]. The value used by Whipple is probably much more appropriate to the photographic meteors, for which it was determined, than for the dust particles in the direct measurements range of sizes.

The effective area of the dust-particle sensor was approximately 0.4 m² to an omnidirectional flux of particles. For particles in a meteor stream, the effective area drops to approximately 0.1 m² because of characteristics of the system. The average influx rate during the 70-hour interval is 2.0×10^{-1} particles m⁻² sec⁻¹ at an average particle mass of 1.4×10^{-9} g. This influx rate is between 1 and 2 orders of magnitude higher than predicted on the basis of the model average distribution established by direct measurements with other satellites and rockets (McCracken and Alexander, to be published).

The manner in which the impact rate varies during the 70-hour interval is indicative of high variability in the spatial density distribution of dust particles in the event. Although corrections for average particle velocity are very important when momentum-sensitive dust-particle sensors are used, the large fluctuations in the impact rate during the November 16-18 event cannot be attributed to a variation in effective mass sensitivity of the system such as would arise from a changing average particle velocity. Neither can the degree of fluctuation in the impact rate be explained on the basis of statistical fluctuations during a series of sampling intervals. The conclusion is reached that the rapid fluctuations in the impact rate are manifestations of real fluctuations in the spatial density of interplanetary dust particles encountered by 1959 Eta on November 16-18, 1959. Since concentrations of dust particles will rapidly be dispersed by the Poynting-Robertson effect and by the cumulative effects of small differences in the orbital elements of the individual dust particles, it seems that the observed concentration of dust particles must have been of rather recent origin.

A quantity that would seem to be of major importance in discussing the presence or absence

small dust particles in a known meteor stream the age of the stream. In line with what is now known about cometary and meteor stream evolution, it seems plausible that the younger streams are more likely to contain small dust particles than are older streams. It seems reasonable also that very young meteor streams are more likely to contain concentrations of dust particles.

Whipple and Wright [1954] investigated the average deviation of photographic meteor radiants from the mean radiant of the meteor stream of various known meteor showers. If either the average deviation of radiants away from the mean radiant or the stream width can be used as a measure of the age of a given stream, one could be led to the conclusion that the Taurids and Perseids are from relatively old meteor streams while the Giacobinids and Leonids are from relatively young meteor streams.

The age of the stream possibly may not be as important an index to dust-particle activity as would be the time elapsed since larger fragments of the stream broke up into dust particles of the sizes observed by the direct measurements technique. Direct measurements of the dust-particle content of the Perseid meteor stream could do so much in helping to answer this question because the Perseid stream is generally regarded as one of the oldest meteor streams.

The 1946 display of the Draconids is usually taken as evidence of a very recent formation of that meteor stream from comet Giacobini-Zinner. High dust-particle activity was not observed by 1959 Eta during the 1959 Draconid meteor shower, which should have been a periodic return of the concentration of meteoroidal debris responsible for the 1946 display. The increase in distance of the stream from earth introduced by Jupiter's perturbation is

probably sufficient to explain the absence of dust particles in the direct measurements range of sizes. Even without the effect of Jupiter's perturbation, a shower would have been unlikely on the basis of the results of Davies and Lovell [1955] showing a relationship between the occurrence of the Draconids and the relative times at which comet Giacobini-Zinner and earth reached the node of the comet's orbit.

In conclusion, the observations made with 1959 Eta have revealed interesting interplanetary dust particle activity at sizes of particles for which ground-based observations are not possible. The impossibility of determining the radiants of the dust particles during the November 16-18, 1959, event demonstrates the need for more sophisticated dust-particle sensors on oriented satellites. It would probably be most desirable to use earth satellites in studying the dust-particle content of meteor streams by the direct measurements technique in order that ground-based observations of meteors could be checked to see what, if any, correlation may exist between the influx of dust particles in widely different ranges of particle mass.

REFERENCES

- Davies, J. G., and A. C. B. Lovell, *MNRAS*, **115**, 23, 1955.
- LaGow, H. E., and W. M. Alexander, Recent direct measurements of cosmic dust in the vicinity of the earth using satellites, in *Space Research*, pp. 1033-1041, North Holland Publishing Company, Amsterdam, 1960.
- Whipple, F. L., *Proceedings VIII International Astronautical Congress, October 6-12, 1957, Barcelona*, Springer-Verlag, Vienna, 1958.
- Whipple, F. L., and F. W. Wright, *MNRAS*, **114**, 229, 1954.

(Received September 1, 1961; revised September 8, 1961.)

Some Micrometeorological Measurements in Ocean Fog

FRANCIS M. WIENER, JAY H. BALL, AND CREIGHTON M. GOGOS

*Bolt Beranek and Newman Inc.
Cambridge, Massachusetts*

Introduction. Audible fog signals are extensively used as aids to navigation during periods of low visibility. Their performance has frequently been less than satisfactory. To gain a better understanding of the propagation of audible sound over ocean waters in fog an experimental study was undertaken which included, in addition to extensive acoustical measurements, a series of micrometeorological measurements in fog. It is well known that the propagation of sound along the earth's surface depends strongly on the vertical gradients of wind and temperature which exist along the transmission path [Rayleigh, 1878; Ingard, 1953a, b; Wiener and Keast, 1959]. The presence of these gradients results in a variation of the speed of sound with height which causes refraction of the sound waves. The program of measurements included, therefore, the determination of the vertical profiles of mean temperature and wind speed in the first 30 feet above the ocean surface. Measurements of the droplet size distribution and liquid water content of fog were also performed, and estimates were made of the horizontal visibility for a variety of fogs.

With the aid of these observations, an attempt was made to account for the experimentally observed values of sound attenuation. These matters are discussed elsewhere [Wiener, 1961]. In this note samples of the observed wind profiles over water in fog are presented, together with data on the corresponding temperature profiles. Such profiles are, to the best of our knowledge, not to be found in the literature. This paper concludes with the presentation of data on liquid water content and droplet size distribution in fog. The influence of the droplet size on visibility, in addition to the liquid water content, is discussed.

Measurements of wind and temperature profiles. The difficulties of obtaining valid measurements of the vertical profiles of mean wind

and temperature over and close to a large body of water are many and varied and are discussed in the recent literature [Deacon, Sheppard, and Webb, 1956; Takahashi, 1958; Wagner, 1958; Portman, 1959]. The presence of fog adds to the problems of instrumentation and maintenance. Although the technique used in the present study has certainly not overcome all difficulties and sources of error, the results have proved useful for the intended purpose, namely, the estimation of the attenuation of sound transmitted over ocean waters in fog. It is hoped that they may be of more general utility also.

The measurements were made on Great Duck Island off the coast of Maine during June 1959. An attempt was made to operate the micrometeorological tower supporting the anemometers and thermocouples [Wiener, Goff, and Keast, 1958] from a specially designed buoy placed in the water offshore. However, the stability of the buoy proved to be inadequate in rough seas, and the advantage of approximately constant instrument heights above sea level under all conditions was not realized. The instrumented tower was then placed on shore near the high-water mark. The portion of the shore around the tower exhibited only a gentle slope, and the tower was placed in such a location that it was exposed directly to the prevailing winds during fog (SW to SE).

Figure 1 shows three wind profiles measured in fog and in fog with light rain. The mean horizontal wind speeds \bar{u} have been normalized to the wind speed \bar{u}_{30} at 30 feet above the water level [Keast and Wiener, 1958], marked on the graph.¹ The sea was choppy, with an

¹ The changes of wind direction with height in the lowest levels of the atmospheric boundary layer are generally small [U. S. Air Force, 1960]. Hence, the profiles were plotted directly in terms of the anemometer readings (magnitude of wind speed in the horizontal plane) in accordance with accepted practice.

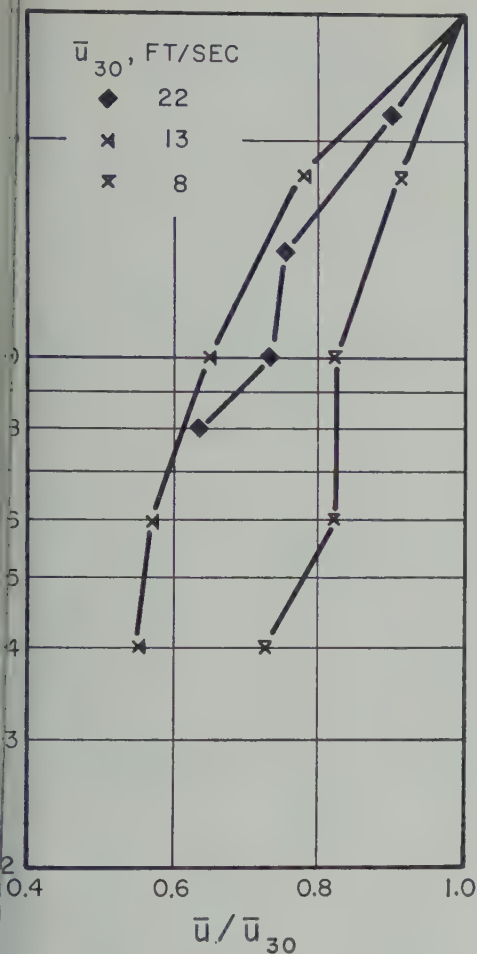


Fig. 1. Samples of mean wind profiles over water in fog and fog with light rain. Temperature stratification is near neutral.

estimated wave height of 2 to 3 ft. The measurements were performed during the late morning and early afternoon as the winds were from the southeast and southwest and as the warm sector of a wave cyclone, moving toward the northeast, had overspread the Great Duck Island area. The modified tropical maritime air mass appears to have come during the previous 48 hours from Pennsylvania and New Jersey State. It arrived at the site with a relatively long over-water fetch. The mean temperature profiles corresponding to two of the profiles of Figure 1 are given in Figure 2. They show a near-neutral temperature stratification, and their shape is very nearly logarithmic with height. This was found to be typical of

advection fog. On the other hand, moderate lapse and inversion conditions ($\pm 2^\circ$ to 3°F mean temperature difference in 30 ft) were sometimes observed in prefrontal and post-frontal fog under transitional conditions.

The data in Figures 1 and 2 are plotted in terms of the approximate instrument heights above mean sea level at the time of observation by considering tide corrections calculated from tide tables adjusted for the locality. The tide corrections for the data in Figures 1 and 2 were small. Examination of other profiles for which this was not the case showed no systematic dependence on the magnitude of the tide correction. This leads one to regard the shore line as a continuation of the ocean surface, but with increased roughness.² The 'kinks' exhibited by two of the three wind profiles shown

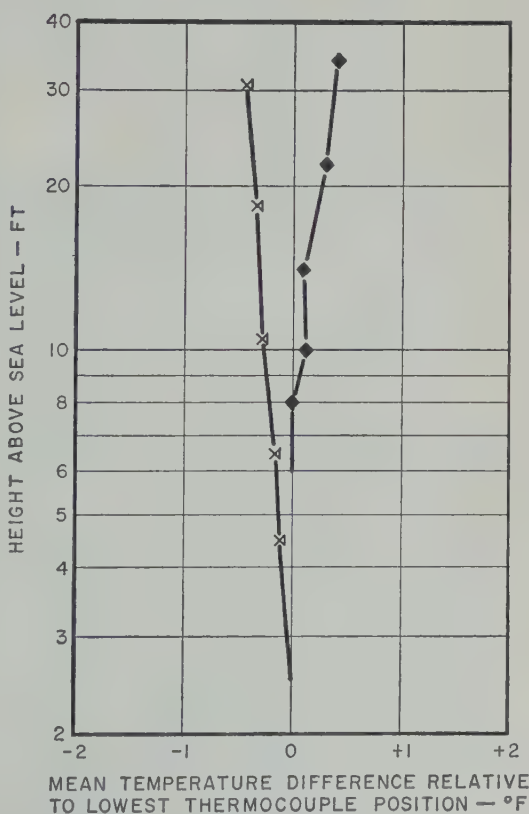


Fig. 2. Samples of mean temperature profiles over water in fog and fog with light rain.

² Observations were only taken during overcast and foggy conditions when the rocks in the vicinity of the tower base were cool and wet.

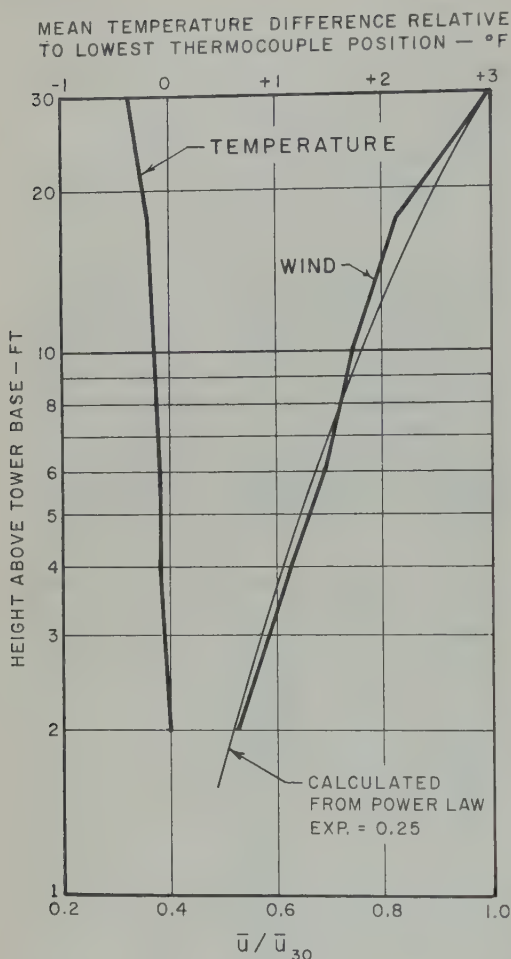


Fig. 3. Average of six mean temperature and wind profiles over water in fog and in fog with light rain. A profile calculated from an assumed power law is shown for comparison.

in Figure 1 may be the result of the increased roughness of the shore. Any such profile change will be only relatively minor for the very short fetches between sea and tower and will be restricted to the lowest layers because the re-adjustment of the wind profile to a marked change in surface roughness generally requires relatively long fetches [Glaser, 1955]. In view of the above, it may be useful to plot all observed profiles in fog in terms of heights from the tower base. This was done in Figure 3 which shows the average of six mean temperature and wind profiles taken in fog and fog with light rain. Examination of Figure 3 shows that the shape of the temperature profiles in

fog is, on the average, approximately logarithmic, as was mentioned above. The shape of the average wind profile in fog can be reasonably well approximated by a power law of the general form $\bar{u} = kz^p$, where k and p are constants and z denotes the height above sea level. The exponent depends, among other things, on the thermal stratification [Davenport, 1960]. In fair weather, values of $p \approx 0.15$ have been calculated from recently reported data obtained over water for near-neutral temperature stratification [Takahashi, 1958]. However, under these conditions, Brooks [1959] reported logarithmic profile shapes. A value of $p = 0.25$ was chosen for the calculated profile shown in Figure 3.

Measurements of the properties of ocean fog. Measurements of droplet size and liquid water content of fog were carried out using the techniques and instrumentation described by Houghton and Radford [1938]. Figure 4 shows a sample of the distribution of the droplet size and the distribution of the liquid water content of the droplets for a fog of liquid water content $LWC = 0.0042 \text{ g/m}^3$. The droplet size carrying the largest percentage of the total liquid water content is seen to be appreciably larger than the droplet size corresponding to the maximum of the size distribution.

Estimates of visibility were obtained by observing a row of marker buoys located in a straight line from the island at accurately known distances.³ To determine visibilities in dense fog a black target $4 \times 7 \text{ ft}$ in size and mounted

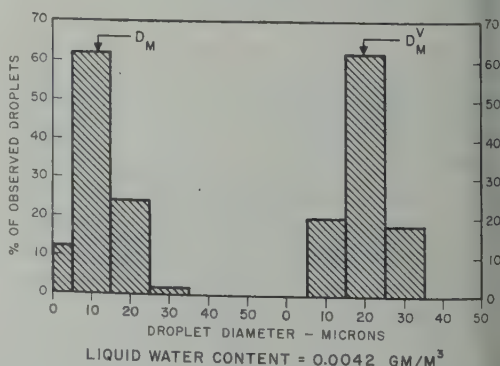


Fig. 4. Distribution of size and water content of droplets in a fog of liquid water content $LWC = 0.0042 \text{ g/m}^3$.

³ These buoys were used to mark the microphone positions when acoustical measurements were made.

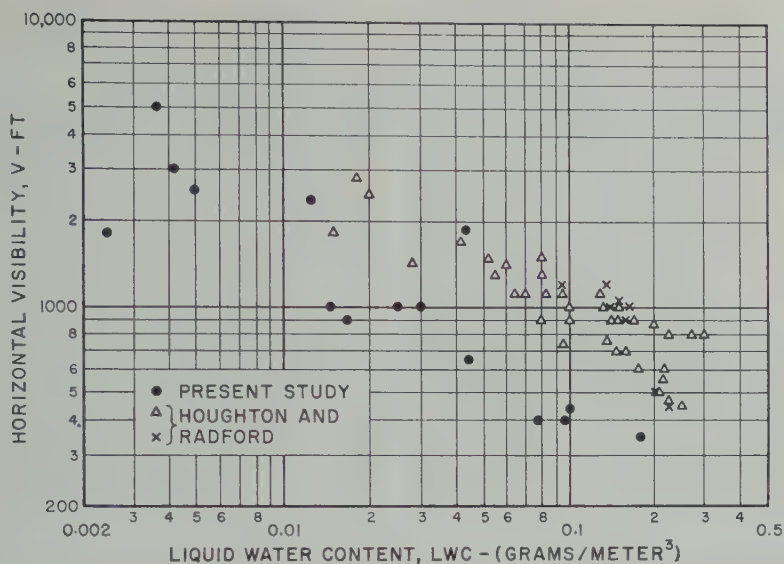


Fig. 5. Horizontal visibility V in ocean fog as a function of liquid water content LWC .

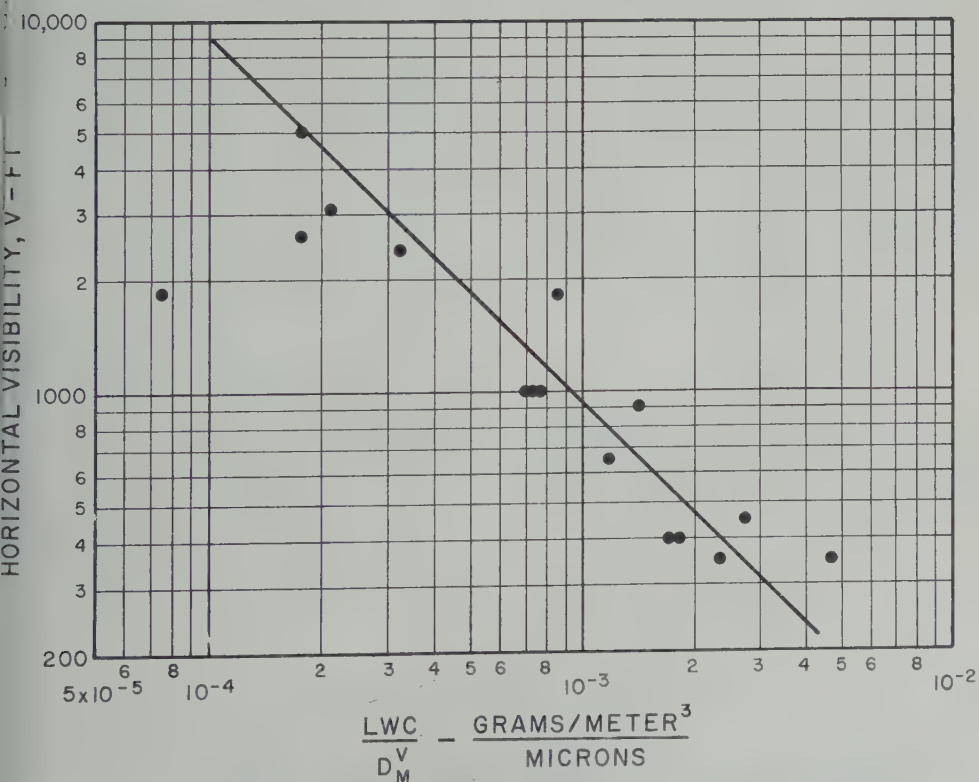


Fig. 6. Horizontal visibility V in ocean fog as a function of liquid water content LWC divided by the droplet side D_M^V carrying the largest percentage of LWC .

on a pole was used. Figure 5 shows a summary of the results of the measurements of horizontal visibility V plotted against the liquid water content of the fog. Houghton and Radford's [1938] results are plotted on the same graph for comparison. It is seen that the scatter of the points is considerable, the points of the present study being characterized by generally lower visibility. This is not unexpected; valid estimates of the visibility are not easily made because the visibility depends also on the size and color of the target and background against which the target is observed. Moreover, visibility in a given fog is seldom constant in time. Nevertheless, the observations suggest that V is not only a function of LWC but also of droplet size, as was observed earlier by Houghton and Radford. The points of Figure 5 have been replotted in terms of the abscissa LWC/D_m^V , where D_m^V denotes the droplet diameter carrying the largest percentage of LWC . The scatter of the experimental points of this study is thereby reduced, and a straight line could be fitted to the data reasonably well, as is shown in Figure 6.

Acknowledgments. We wish to extend our sincere thanks to the many persons whose efforts made this difficult measurement program possible. In particular, our most sincere appreciation is extended to the officers and men of the U. S. Coast Guard concerned with this project. Thanks are also due Dr. H. E. Cramer of the Massachusetts Institute of Technology for his assistance in interpreting the data. This work was supported by the U. S. Coast Guard.

REFERENCES

- Brocks, K., Measurements of wind profiles over the sea and the drag at the sea surface, Preprints, International Oceanographic Congress, August-September 1959.
- Davenport, H. G., Rationale for determining design wind velocities, *J. Struct. Div., Proc. A Soc. Civil Engrs.*, 86, ST5, 39-68, 1960.
- Deacon, E. L., P. A. Sheppard, and E. K. Wells, Wind profiles over the sea and the drag at the sea surface, *Australian J. Phys.*, 10, 511-541, 1957.
- Glaser, A. H., The temperature above an airport runway on a hot day: moist climate, *Sci. Rep.* 4, Texas A & M Research Foundation, 1955.
- Houghton, H. G., and W. H. Radford, On the measurement of drop size and liquid water content in fogs and clouds, *Papers Phys. Oceanogr. Meteorol., Mass. Inst. Technol. and Woods Hole Oceanog. Inst.* 6 (4), 1938.
- Ingard, Uno, The physics of outdoor sound, *Pr. Fourth Ann. Natl. Noise Abatement Symposium* 4, 11-25, October 1953a.
- Ingard, Uno, Review of the influence of meteorological conditions on sound propagation, *Acoust. Soc. Am.* 25, 405-411, 1953b.
- Keast, D. N., and F. M. Wiener, An empirical method for estimating wind profiles over open level ground, *Trans. Am. Geophys. Union*, 39, 858-864, 1958.
- Portman, D. J., An improved technique for measuring wind and temperature profiles over water and some results obtained for light winds, *Preprints Third Conf. Great Lakes Research*, pp. 77-84, May 1959.
- Rayleigh, John William Strutt, 3d baron, *The Theory of Sound*, 1878, reprinted by Macmillan and Co., London, vol. 2, pp. 132-138, 1940.
- Takahashi, T., Micro-meteorological observations and studies over the sea, *Mem. Fac. Fisheries Kagoshima Univ.*, 6, 158 pp., 1958.
- U. S. Air Force, *Handbook of Geophysics*, revised ed., Macmillan and Co., New York, chap. 1, 1960.
- Wagner, N. K., An analysis of some over-water wind profile measurements, *Trans. Am. Geophys. Union*, 39, 845-852, 1958.
- Wiener, F. M., Sound propagation over oceans and waters in fog, *J. Acoust. Soc. Am.*, 33, Sept. 1960.
- Wiener, F. M., K. W. Goff, and D. N. Keast, Instrumentation for study of propagation of sound over ground, *J. Acoust. Soc. Am.*, 30, 860-864, 1958.
- Wiener, F. M., and D. N. Keast, Experimental study of the propagation of sound over ground, *J. Acoust. Soc. Am.*, 31, 724-733, 1959.

(Received April 6, 1961; revised August 28, 1961)

Reply to Dr. Harold C. Urey's Criticism of the Paper by Brian Mason, 'The Origin of Meteorites'

BRIAN MASON

*The American Museum of Natural History
New York 24, New York*

The principal point of issue between Dr. Urey and myself seems to be whether the carbonaceous chondrites represent a source material from which the 'common' olivine-pyroxene chondrites have been formed by dehydration and partial reduction at temperatures above 600°C, as I suggested. Dr. Urey believes this is impossible and suggests that "some water containing compounds, hydrogen sulphide, etc., infiltrated some high iron group chondritic material, oxidized the metallic iron to magnetite and the sulphide, deposited carbon compounds, sulfate, sulphur, etc., and removed some sodium and potassium, and partly destroyed the chondrites," thereby producing the carbonaceous chondrites.

Dr. Urey evidently held opinions similar to my own some years ago, but rejected them because the carbonaceous chondrites and the olivine-pyroxene chondrites do not have constant composition with respect to the nonvolatile elements. Nevertheless, the correspondence is close, as Dr. Urey himself admits when he writes, "It is difficult in some cases to decide as to whether the variations (between the two groups) are real or due to analytical error." He points out differences between the percentages of Na and K and the Mg/Si and Fe/Si ratios for carbonaceous chondrites and other types of chondrites. It is doubtful how much weight should be given to these differences. Apart from possible analytical error (probably significant in view of the extremely difficult problems involved in the analysis of carbonaceous chondrites), it should be borne in mind that the carbonaceous chondrites contain water-soluble compounds of sodium, potassium, and magnesium, and hence these elements may readily be lost, as pointed out by Ringwood [1961, p. 170].

In his criticism of my paper Dr. Urey postulates a constant composition for the material from which the meteorites were formed, and a closed system throughout. There seems to be no

factual evidence for either postulate. Indeed, variations of physical conditions in different parts of the primordial solar system should be reflected in variations of chemical composition, and the solar system or parts thereof is probably not completely closed in the physicochemical sense. It is quite remarkable that carbonaceous chondrites, especially within the individual types established by Wiik, are so uniform in composition, especially in view of the presence of volatiles, water-soluble compounds, and amorphous material.

Dr. Urey places great weight on the nature of the metal particles in chondritic meteorites—particles which he describes as fragments from some pre-existing bodies. It is surely remarkable that the composition of these so-called fragments is directly related to the bulk composition of the meteorite in which they occur, a relationship first established by Prior and confirmed by other investigators [Ringwood, 1961]. Such a relationship indicates a chemical equilibrium, not a chance aggregate of fragments.

Dr. Urey states, "Heating the carbonaceous chondrites in a vacuum is difficult because they are dispersed quite easily out of a crucible." This is not so: I have heated fragments of Cold Bokkeveld and Murray in a vacuum at about 1000°C—they did not disintegrate but shrunk slightly, and the resultant product gave an X-ray diffraction pattern of olivine.

However, it is difficult to produce decisive evidence for a clear-cut decision between Dr. Urey's hypothesis and my own for the relationship between carbonaceous chondrites and olivine-pyroxene chondrites, since we differ fundamentally as to whether the olivine-pyroxene chondrites were derived from the carbonaceous chondrites, or vice versa; arguments based on chemical composition alone cannot resolve this conflict of opinion because the postulated chemical reactions may go in either direction,

according to conditions of temperature, pressure, and chemical environment. However, if the carbonaceous chondrites were formed as envisaged by Professor Urey one would expect to find in them chondrules of serpentine pseudomorphous after olivine or pyroxene. I have searched for serpentine chondrules in carbonaceous chondrites and not found them; *Kvasha* [1948] described chlorite chondrites in the Boriskino carbonaceous chondrite, but after studying her description and the photographs I believe that these chondrules were actually of iron-free olivine, which has low birefringence and refractive index within the range of chlorite. The chondrules in carbonaceous chondrites are of olivine, or occasionally of pyroxene, and appear to have replaced the serpentine, which forms the matrix. In spite of the presence of much ferrous iron, the olivine in the Type II carbonaceous chondrites is almost or completely iron-free, which is characteristic of olivine formed by the thermal decomposition of serpentine [*van Biljon*, 1960]. In Type I carbonaceous chondrites the silicate is amorphous, and they contain no chondrules. Thus the sequence appears to be: Type I carbonaceous chondrites (no chondrules) → Type II carbonaceous chondrites (chondrules forming by decomposition of serpentine) → Type III carbonaceous chondrites (serpentine completely converted to olivine + pyroxene), rather than the reverse.

I would like to take this opportunity to correct an erroneous statement in my original paper. On p. 2967 the sentence "X-ray powder photographs of Alais, Cold Bokkeveld, Haripura, Mighei, Murray, Nawapali, Nogoya, and Santa Cruz show that these are also largely chlorite, or a related mineral with the structure of serpentine; powder photographs of Orgueil, Ivuna and Kaba show lines of magnetite only." This sentence should read "X-ray powder photographs of Cold Bokkeveld, Haripura, Mighei, Murray, Nawapali, Nogoya, and Santa Cruz show that these are also largely chlorite, or a related mineral with the structure of serpentine; powder photographs of Orgueil, Ivuna, and Alais show lines of magnetite only."

REFERENCES

- Kvasha*, L. G., Investigation of the stony meteorite Staroye Boriskino, *Meteoritika*, 4, 83-96, 1948.
Mason, Brian, The origin of meteorites, *J. Geophys. Research*, 65, 2965-2970, 1960.
Ringwood, A. E., Chemical and genetic relationships among meteorites, *Geochim. et Cosmochim. Acta*, 24, 159-197, 1961.
Urey, H. C., Criticism of Dr. B. Mason's paper on "The origin of meteorites," *J. Geophys. Research*, 66, 1988-1991, 1961.
van Biljon, W. J., Chemical reactions in the thermal breakdown of the serpentine minerals, *Neues Jahrb. Mineral. Abhandl.*, 94, 1223-1240, 1960.

(Received September 14, 1961.)

A Note on the Thermodynamic Theory of Nonhydrostatically Stressed Solids

MINEO KUMAZAWA

*Institute of Earth Sciences, Nagoya University
Nagoya, Japan*

In a recent paper printed in this journal, Kamb [1961] reviewed the thermodynamic theory of nonhydrostatically stressed solids. I agree with his view cited as follows: 'Since the Gibbs free energy is only a secondarily defined quantity in ordinary hydrostatic thermodynamics, energy and entropy being the primary quantities of physical significance, it is of course not permissible in nonhydrostatic thermodynamics to attempt to define a Gibbs free energy in a suitable way' (p. 265). However, I should like to comment on his discussion.

Kamb presented the theoretical basis both for preferred orientation of minerals and for phase transformation under the nonhydrostatic stress (Appendix 2, p. 269). His derivation of the equilibrium condition does not seem to be explicit. His equation [49] (hereafter equations that appeared in Kamb's [1961] paper are denoted by []) must be:

$$\begin{aligned} &= T \delta S_s - PA(\delta l + \delta x) \\ &+ \epsilon \frac{A \delta x}{V} + T\eta \frac{A \delta x}{V} \\ &+ T \delta S_F - PA \delta l \\ &+ \mu \delta n \left[+ T\eta' \frac{A}{V} \delta n \right] \quad (1) \end{aligned}$$

where η' is the partial entropy of the solid component in the fluid phase. The second term represents the work done on the solid phase through the displacement $\delta l + \delta x$ (not δl) of the interface between solid and fluid phases. The term in brackets is either necessary or unnecessary according to the definition of S_F . Equation [] and $\delta S = 0$ imply that the equilibrium state is specified under the adiabatic condition. Equation [51] represents the isolated system which does not give the work on the surroundings. Thus, the variation of total internal energy is

minimum or vanishes for the equilibrium, and we obtain

$$\delta E = \left(\frac{\epsilon}{V} - \frac{\mu'}{V} \right) A \delta x = 0 \quad (2)$$

or

$$\epsilon = \mu' \left[= \left(\frac{\partial E_F}{\partial n} \right)_{S_F, V_F} \right] \quad (3)$$

This relation is nothing but the balance of the adiabatic and isovolumetric thermodynamic functions of both phases (i.e. the partial internal energy of the solid component in the fluid μ' is equal to the internal energy of the solid phase ϵ). Under the isobaric ($\delta P = 0$) and isothermal ($\delta T = 0$) conditions, we have his equation [1] or [53]:

$$\epsilon - T\eta + PV = \mu \left[= \left(\frac{\partial G_F}{\partial n} \right)_{T, P} \right] \quad (4)$$

where the terms on the left side are the quantities in the solid phase, and μ is a chemical potential or partial Gibbs free energy of the solid component dissolved in the fluid phase. The quantity $\epsilon - T\eta + PV$ is determined by the environmental conditions P and T and also by the equation of state expressing the physical properties of the solid itself. Consequently, Kamb carried out the calculation of $\epsilon - T\eta + PV$ by use of the elastic constants of the solid and discussed the development of preferred orientation of crystallographic axis of anisotropic minerals under stress. However, he states that 'the Gibbs-Kamb theory does not give a Gibbs free energy for the solid, but only for the component dissolved in the fluids adjacent to the solid; ...' (p. 261). This argument contradicts his method of practical calculation of the quantity $\epsilon - T\eta + PV$, and it seems to me that his fundamental idea of thermodynamics is inadequate.

In addition, he commented on *MacDonald's* [1960] theory in which the variation of strain energy or Helmholtz free energy

$$\delta F = \frac{1}{2} \int_{\tau} \sigma_{ij} \delta \epsilon_{ij} d\tau \quad (5)$$

was derived from

$$F = \frac{1}{2} \int_{\tau} \sigma_{ij} \epsilon_{ij} d\tau \quad (6)$$

Kamb stated that the quantities ϵ_{ij} in (5) are not the same as those in (6) and this difference was caused by the rotation of reference frame. If the deformation and rotation are finite, F cannot be expressed by (5) or (6) and must be always written as

$$\delta F = \int_{\tau} \left[\sigma_{ij} \frac{\partial \delta u_i}{\partial i} + \sigma_{ij} \left(\frac{\partial \delta u_j}{\partial i} + \frac{\partial \delta u_i}{\partial j} \right) \right] d\tau \quad (7)$$

where u_i are the components of virtual displacement. Four kinds of definition of strain have been presented [Shimazu, 1954]. However, (7) cannot be written by a linear function of the variation of strain components of any kind. If we write the strain energy by a linear function of strain components, the infinitesimal deformation is implicitly assumed and the effect of rotation of the reference frame is excluded. Under the infinitesimal condition we have

$$\delta F = \int_{\tau} \sigma_{ij} \delta \epsilon_{ij} d\tau \quad (8)$$

Integrating (8) under the assumption of linear elasticity, we have (6). In the special case in which σ_{ij} are kept constant throughout the deformation, we have

$$F = \int_{\tau} \sigma_{ij} \epsilon_{ij} d\tau \quad (9)$$

The variation of (6) or (9) is reduced to (8), which is different from (5), regardless of the finite or infinitesimal deformation and rotation.

Recently I also investigated the same problem as Kamb's and obtained results different from both those of Kamb and of others he referred to. My results are presented here briefly. Combining the first and the second laws of thermodynamics, we may define the stability condition

by the theorem that a certain thermodynamic function is minimum according to the environmental condition specifying the system. The general form of the thermodynamic function ϕ is given by

$$\phi = U - X_i x_i \quad (10)$$

or

$$\delta \phi = X_i \delta x_i - x_i \delta X_i, \quad (i \neq j)$$

where X are intensive quantities: T , P_{xx} , P_{yy} , P_{zz} , P_{yz} , P_{zx} and P_{xy} ; and x are the extensive quantities: S , V_{xx} , V_{yy} , V_{zz} , V_{yz} , V_{zx} and V_{xy} . P_{ij} and V_{ij} are the generalized pressure and volume, respectively. The components of generalized pressure are given directly by those of the stress tensor as follows:

$$P_{ij} = \sigma_{ij}$$

The components of volume are given by

$$V_{ij} = V_0 \epsilon_{ij} + C_{ij}$$

or

$$\delta V_{ij} = V_0 \delta \epsilon_{ij}$$

where V_0 , ϵ_{ij} , and C_{ij} are the volume in the ordinary sense, the components of infinitesimal strain tensor, and the constants expressing the shape or size of the system, respectively.

In conclusion we have 27 kinds of thermodynamic functions in the nonhydrostatic thermodynamics, while in the ordinary thermodynamics we have only 4 kinds of thermodynamic functions (internal energy, enthalpy, Helmholtz free energy, and Gibbs free energy). The quantity C_{ij} is equal to $V_0 \Delta_{ij}/3$ for a system in the form of a cube, where Δ_{ij} is the Kronecker delta. In the case of a system of rectangular parallelepiped, C_{ij} is given approximately by

$$C_{ij} = [V_0 \Delta_{ij}/3][(3l_i - 2V_0^{1/3})/V_0^{1/3}]$$

where l_i is the edge length of the system along the direction of the coordinate axis i . Therefore V_{ij} are the state variables expressing both the size and the shape of the system. The exact determination of V_{ij} must be derived according to the theory of finite deformation.

Under the isobaric and isothermal condition (10) is reduced to 'the generalized Gibbs function is minimum' or

$$\delta G = -S \delta T - V_{ij} \delta P_{ij} = 0 \quad (11)$$

generalized Gibbs function for a cubic system is identical with the Gibbs free energy by MacDonald [1957]. The Gibbs function defined here is dependent not only upon the shape but also upon the shape of the system, it is independent of the orientation and is uniform throughout the system if the stress and strain are uniform. Under the isothermal and isovolumetric condition (10) is expressed by 'the generalized Helmholtz free energy is minimum'

$$\delta F = -S \delta T + P_{ij} \delta V_{ij} = 0 \quad (12)$$

which is identical with the minimum condition for common Helmholtz free energy.

In ordinary thermodynamics, the term 'isostatic' means that the pressure or equivalent hydrostatic pressure of the system is maintained constant during changes of the system. Under nonhydrostatic condition, however, the term 'isostatic' must mean that all the components of stress, i.e. the components of the stress tensor, are maintained constant, since we have independent state variables P_{ij} instead of $-\frac{1}{3}\Sigma P_{ij}\Delta_{ij}$. Likewise, the term 'isovolumetric' means here that six state variables V_{ij} are all maintained constant throughout any changes. In a general case of an equilibrium, the maintenance of isobaric or isovolumetric can hardly be maintained. When two phases contact each other in a plane under a given nonhydrostatic stress field, for instance, the normal pressure is maintained constant while the lateral one is not. Environmental physical conditions specifying equilibrium are complicated, especially in a heterogeneous polycrystalline aggregate such as a mineral. Therefore I investigated the intermediate conditions between (11) and (12), and the continuity conditions of stress and strain at the boundary are considered as the first approximation. I introduced a new thermodynamic function ϕ_s (function, or ϕ_s potential)

$$\phi_s = U - TS - \sum_{i=x,y,z} V_{is} P_{is} \quad (13)$$

$$\begin{aligned} \delta \phi_s = & -S \delta T + \sum_{i=x,y,z} P_{ij} \delta V_{ij} \\ & - \sum_{i=x,y,z} V_{is} \delta P_{is} \end{aligned}$$

where the Z axis of the coordinate is taken as normal to the plane boundary separating the

two phases. The minimum condition of the ϕ_s function is applicable to the system which is isobaric and volumetric with respect to the pressure or stress (P_{xx} , P_{yy} , and P_{zz}) and the volume (V_{xx} , V_{yy} , and V_{zz}), respectively. The situation of environmental condition mentioned above is, in some respects, analogous to that presented by Kamb in Figure 1 (p. 269), where one-dimensional equilibrium is concerned. The thermodynamical potential defined by the ϕ_s function is thus dependent not only on the shape and size of the system but also on the orientation. A ϕ_s function may be called the chemical potential, since it defines a driving force of the chemical reaction under nonhydrostatic stress field. A ϕ_s function is uniform throughout the uniformly stressed or strained body, although it is different on each surface of the body.

I have reached the following conclusions. When a nonhydrostatic stress is given, the stability condition is specified by the maximum strain energy (minimum of Gibbs free energy). On the other hand, if a nonhydrostatic condition is given by the strain, the stability is specified by the minimum strain energy (minimum of Helmholtz free energy). Therefore, a state of the maximum or minimum of strain energy is not a sufficient condition to bring about a stable phase (or crystallographic orientation). The criterion of stability specified by the ϕ_s function is the elastic softness, which is identical with neither the simple compliance constant nor the other elastic parameters given by the previous investigators.

Based upon the generalized thermodynamic function (10), it is concluded that the equilibrium state under the nonhydrostatic condition is specified not only by the mean hydrostatic pressure but also by the anisotropy of the stress field as well as by the mechanical structure of the system. Consequently, the effect of shear stress upon the chemical reaction is not negligible, contrary to the generally accepted view. For instance, the Clapeyron-Clausius equation for the first-order phase equilibrium is given by

$$dT/dP_s = -\Delta V/\Delta S$$

where P_s is the normal pressure (extensional) on the plane between the two phases at equilibrium, but is not an equivalent hydrostatic pressure. Let us consider a mineral which is stable under

the hydrostatic pressure P_0 . This mineral can be stable along the plane, provided the normal pressure on the plane is higher than P_0 , even if the equivalent hydrostatic pressure is lower than P_0 . The conclusion noted above is applicable to the discussion of sharpness of the discontinuity within the earth, e.g. the basalt-eclogite phase transition hypothesis on the Mohorovicic discontinuity. It will be also applicable to the studies of physical conditions prevailing in the metamorphic facies.

The details will be reported in the Journal of the Institute of Earth Sciences, Nagoya University.

Acknowledgments. I should like to express my hearty thanks to Prof. Kumiji Iida for his guid-

ances and encouragement. My thanks are also due to Dr. Yasuo Shimazu, who gave valuable suggestions to me.

REFERENCES

- Kamb, W. B., The thermodynamic theory of non-hydrostatically stressed solids, *J. Geophys. Res.*, **66**, 259-271, 1961.
- MacDonald, G. J. F., Thermodynamics of solids under non-hydrostatic stress with geologic applications, *Am. J. Sci.*, **255**, 266-281, 1957.
- MacDonald, G. J. F., Orientation of anisotropic minerals in a stress field, *Geol. Soc. Am. Mem.*, **79**, 1-8, 1960.
- Shimazu, Y., Equation of state of materials composing the earth's interior, *J. Earth Sci., Nagoya Univ.*, **2**, 15-172, 1954.

(Received June 22, 1961.)

Author's Reply to Discussions of the Paper
'The Thermodynamic Theory of Nonhydrostatically Stressed Solids'

W. BARCLAY KAMB

California Institute of Technology
Pasadena, California

the comments of MacDonald [1961], Hoffer [1961], and Kumazawa [1961] on the recent paper of mine [Kamb, 1961] call attention to some aspects of the thermodynamic theory of nonhydrostatically stressed solids that are worth discussing.

In reiterating his earlier statement [MacDonald, 1960, p. 4] that the equilibrium orientation of a crystal in a nonhydrostatic stress field does not depend on the mechanism by which the orientation is determined, MacDonald is making an assertion that in my opinion requires proof which has not been given. The same assertion is made, equally categorically, by Hoffer [1961]. He fails to distinguish between the thermodynamic theory *per se* (the subject of my recent paper) and its application, with the introduction of further assumptions, to derivation of a fabric (the subject of an earlier paper [Kamb, 1959]). Certainly the thermodynamic condition of equilibrium—in the case I have considered, for equilibrium between a nonhydrostatically stressed solid and an adjacent fluid—must be independent both of the manner of derivation of the result and of the process by which the imputed equilibrium is reached. Indeed the point of writing my recent paper was to do with the untenable situation in which there appeared to be two different valid thermodynamic derivations leading to different conditions of equilibrium, ostensibly the same state of equilibrium. From the required uniqueness of the thermodynamic condition for equilibrium of a specific system it does not necessarily follow that different possible crystal-reorienting mechanisms must lead to the same final crystal orientation. The mechanism discussed in my 1959 paper, that of solution and redeposition, involves the irreversible transfer of material from place to place in polycrystalline material. It is inherent in Hoffer's [1906] theory that true equilibrium can

be at best only local in a polycrystalline aggregate under nonhydrostatic stress, and that, if diffusion or other irreversible transfer processes can operate at a finite rate under the prevailing conditions, true general equilibrium will never be attained until the state of stress is reduced to a hydrostatic one. Thus in my fabric theory, even if and when all the crystals of a given mineral species were to reach their most preferred orientation, the aggregate as a whole would not have reached true general equilibrium and the component transfer would continue to take place, as long as the stress is maintained nonhydrostatic. Under such conditions it is obviously impossible to try to assign, in Hoffer's language, 'a given state of free energy' to the system and to expect that this will be minimized when the crystals reach a certain 'equilibrium' orientation, because equilibrium in the true sense does not prevail, and we can at most use equilibrium thermodynamics to describe the locally varying states of near-equilibrium that may be attained locally if the irreversible transfer processes operate slowly enough. Thus there is no *a priori* guarantee, and in fact it seems unlikely, that one can speak properly of an *equilibrium* orientation of a crystal species in a nonhydrostatic stress field, in the way that MacDonald does. When we turn our attention to other possible crystal-reorienting processes—dislocation motion, grain boundary migration, direct granular rotation (as in a loose granular aggregate)—where the physical laws operating are quite different ones from those governing solution and redeposition, it is simply too much to expect that the former any more than the latter will be describable by a general thermodynamic energy function that will necessarily lead in every case to the same final orientation.

It is perhaps true in some cases that if one considers the totality of all possible crystal-re-

orienting processes, acting for a time indefinitely long, there will be a certain final orientation for a given crystal species in a completely specified physical system. The qualification 'perhaps' must be retained because one can imagine *a priori* the possibility that two competing processes, such as recrystallization and intracrystalline creep (gliding or twinning), could go on at essentially equal rates and could each by itself lead to a distinctly different preferred orientation, so that at any time, no matter how remote, there would always be found newly grown (recrystallized) crystals in one orientation and older crystals in different orientations modified from the first by the processes of gliding and/or twinning. When two (or more) irreversible processes compete in producing crystal reorientation, and if each one acting by itself tends to lead to a different orientation, the ultimate orientation or distribution of orientations that result must be determined by the relative rates of the different processes. A simple and striking example of such a situation is provided by the intracrystalline plastic flow of zinc [Rosi, Dube, and Alexander, 1953] in which gliding leads to an orientation favorable to twinning and the resulting twinning to an orientation again favorable for gliding, so that the ultimate distribution of crystal orientations is determined by the relative rates at which the individual zinc crystals pass through the separate stages of twinning and gliding.

It is for such reasons that I decline to accept without proof MacDonald's and Hoffer's assertion that the actual physical processes are not determinative of the preferred orientation, and also that I doubt that a proof of this will be forthcoming.

One should not attribute to me the statement that MacDonald's theory describes the results of certain idealized experiments. MacDonald has not shown any connection between his mathematical treatment and any actual physical process or experiments, relying instead on the general principle of an equilibrium state discussed above. The idealized experiment to which MacDonald's theory appeared to me at first sight to be applicable is that of the crystalline sphere discussed in my paper, but my analysis shows that the theory does not in fact correctly describe such experiments.

In commenting on my discussion of the formulation of the virtual strain, MacDonald refers

to an infinitesimal change in the stress acting on the crystal, whereas in his own treatment that I am discussing, the stress is specifically held constant and only an infinitesimal change in orientation (an infinitesimal rotation) is considered. In this latter case, it is simply untrue that the contributions from rotation to the 'apparent' virtual strain in a fixed coordinate system' can be neglected with respect to those from actual deformation of the crystal. The proof of this is in Appendix 1 of my paper. If a variation of stress is admitted (which corresponds to the general elastic case treated by Sokolinkoff [1956]) it is obviously no longer permissible to use MacDonald's expression $\frac{1}{2} \tau_{ij} \delta e_{ij}$ (tensor notation) for the virtual strain, which presupposes a constant τ_{ij} . The situation under consideration is in no way related to that of the thin elastic rod and the rotations as well as deformations I have discussed are all infinitesimal, not finite.

It would be helpful to have a more complete listing of the inconsistencies that MacDonald reports to find in my paper, because the one he does mention allows illuminating comment. MacDonald finds it unacceptable that a scalar quantity, crystal specific volume, can be a function of crystal orientation in a nonhydrostatic stress field, although he does not indicate any error in the underlying equations (8) and (11). The difficulty here is a misunderstanding of the concept of tensor invariance. The fact that the scalar quantity e_{kk} is invariant with respect to a mathematical rotation of the axes of the coordinate system in which it is calculated does not require that for a given crystal the quantity must remain constant under an actual physical rotation of the crystal. From the relevant equations it is easy to convince oneself that for elastically anisotropic crystals (except isometric ones) the quantity e_{kk} , and hence the specific volume, does vary with crystal orientation (as contrasted with coordinate axis orientation) in a nonhydrostatic stress field, which after all what physical intuition tells us, since the specific volume of a crystal elastically deformed in a particular direction will be decreased when this weak direction is aligned with the direction of the greatest compressive stress. The elastic strain energy $\frac{1}{2} \tau_{ij} e_{ij}$ also varies with orientation of an anisotropic crystal under nonhydrostatic stress, a fact which does not seem to bother MacDonald (it is the main ingred-

s own theory) even though this quantity is a scalar and is invariant with respect to a rotation of the coordinate system.

I do not claim that my theory of preferred crystal orientation [Kamb, 1959] gives results that have been substantiated by experiment or derivation. I believe that it is correctly derived from the assumptions and the physical model which it is based, which are open to question by anyone. I believe that MacDonald's [1960] paper contains invalidating errors and that there is therefore no point in comparing its predictions with experiment. If experiment should show that maximum elastic strain energy is in fact the quantity that correctly determines preferred crystal orientation, as has already been pointed out by Sonder [1933, 1948], it will be for reasons other than those advanced by MacDonald.

The rest of Hoffer's [1961] comments, aside from the point already discussed, are actually an attack on my earlier paper [Kamb, 1959] rather than on the recent paper in this journal, and they are as much an attack on MacDonald's theory as on my own. I think that the difficulties in applying such theories in practice are not as formidable as Hoffer depicts. In any case the proper scientific question is whether such theories are right or not—whether they are soundly based and whether they properly describe the phenomena involved—not whether or not they happen to be easy to apply to specific petrologic problems. I find it amusingly inconsistent that Hoffer objects on the one hand to the complexity of applying even a first-approximation fabric theory of the kind I have given, and on the other hand that the theory predicts not only the general tendency to preferred orientation and the most probable orientation but also exactly the degree of randomness to be expected in fabrics produced under given conditions. It also seems inconsistent to me to insist on the one hand that all crystal-reorienting processes must necessarily lead to the same preferred orientation and then to object on the other that effects that may well be secondary—lack of original random orientation, presence of compositional zoning, polymodal particle sizes, particle shape, variation of the stress field, or failure to attain completely a 'steady state' (I suppose Hoffer here means actually *equilibrium*)—may completely vitiate any attempt to analyze physi-

cally the development of preferred orientation.

It is regrettable that Kumazawa [1961], in commenting on my paper, and in particular on Appendix 2 therein, has not also studied and commented upon Gibbs' [1906] original work. As I have pointed out, my derivation of Gibbs' result is nothing but a simplification of Gibbs' treatment, enabling the underlying physical principles to be more easily appreciated but not changing the basic content of Gibbs' derivation. Thus the faults that Kumazawa finds in my derivation must equally be faults in Gibbs' original work. Actually, I am afraid the errors are Kumazawa's. The main errors are the use of the expression $\delta l + \delta x$ in the second term of his equation 1 in the place where I (and Gibbs) have used only δl , and the inclusion of the additional term $T\eta\delta x/V$ in this equation. Clearly the work done on the solid *originally present* is only $-p\delta l$. In calculating the contribution of the newly crystallized solid layer to the total change in energy of the system one cannot include a 'work term' $-p\delta x$ and a 'heat term' $T\eta\delta x/V$ because the term already written, $\epsilon\delta x/V$, designates all the energy contained in the newly crystallized solid layer, from whatever source it may be derived. Such additional terms would enter only if we wrote the change in energy E_L of the infinitesimal layer in the form

$$\delta E_L = \left(\frac{\partial E_L}{\partial n_L} \right)_{V_L, S_L} \delta n_L + \left(\frac{\partial E_L}{\partial S_L} \right)_{n_L, V_L} \delta S_L + \left(\frac{\partial E_L}{\partial V_L} \right)_{n_L, S_L} \delta V_L$$

which we have not done, for reasons that become obvious if one tries to evaluate the terms in this expression (note that in any case the 'heat term' would *not* be $T\eta\delta x/V$, as given by Kumazawa). We have written instead

$$\delta E_L = \epsilon \delta n_L$$

which represents the definition of the specific energy ϵ .

Kumazawa is led to an equilibrium requirement (or condition) under variation at constant volume and entropy that is different from the requirement under variation at constant pressure and temperature. This conclusion is fundamentally erroneous, equally well for hydrostatic systems as for nonhydrostatic ones, since

equilibrium is a *condition* of the system and the requirements for it do not depend on whether a variational principle for constant S and V or for constant p and T is employed. The formulation of these different but equivalent variational descriptions of (or requirements for) equilibrium is well discussed in Gibbs' original work, which is worthy of restudy in this connection. As mentioned in my paper, if the derivation is carried out correctly the same result is obtained for either type of variational treatment.

The last term, in parentheses, in Kumazawa's equation 1, which does not belong there with the definitions I have employed, does not seem to show up in his other equations and does not need to be discussed further.

There is no contradiction in calculating from the properties of the solid the chemical potential of the dissolved solid component that must be in equilibrium with it. This is what is always done in conventional hydrostatic thermodynamics. If one wishes to associate the chemical potential so calculated with the solid too, one is at liberty to do so, but for nonhydrostatically stressed solids there is no single such value and it is therefore impossible to speak of the chemical potential of a nonhydrostatically stressed solid. It is for this reason that I think the concept loses its usefulness.

The remainder of Kumazawa's discussion, which seems to proceed from the idea of associating a variety of chemical potentials with a nonhydrostatically stressed solid, as a function of the interface across which equilibrium takes place, contains no derivations or proofs that can be examined critically and I would want to

withhold judgment on it until these necessary derivations and proofs are presented.

REFERENCES

- Gibbs, J. W., On the equilibrium of heterogeneous substances, in *Collected Works of J. Willard Gibbs*, Yale University Press, New Haven, Conn., 1906.
- Hoffer, A., Discussion of Paper by W. Barclay Kamb, 'The thermodynamic theory of nonhydrostatically stressed solids, *J. Geophys. Research*, **66**, 2600, 1961.
- Kamb, W. B., Theory of preferred crystal orientation developed by crystallization under stress, *J. Geol.*, **67**, 153-170, 1959.
- Kamb, W. B., The thermodynamic theory of nonhydrostatically stressed solids, *J. Geophys. Research*, **66**, 259-271, 1961.
- Kumazawa, M., A note on the thermodynamic theory of nonhydrostatically stressed solids, *J. Geophys. Research*, **66**, 3981-3984, 1961.
- MacDonald, G. J. F., Orientation of anisotropic minerals in a stress field, *Geol. Soc. Am. Mem.*, **79**, 1-8, 1960.
- MacDonald, G. J. F., Discussion of paper by W. Barclay Kamb, 'The thermodynamic theory of nonhydrostatically stressed solids,' *J. Geophys. Research*, **66**, 2599, 1961.
- Rosi, F. D., C. A. Dube, and B. H. Alexander, Mechanism of plastic flow in titanium, *Trans. AIME*, **197**, 257-265, 1953.
- Sokolnikoff, I. S., *Mathematical Theory of Elasticity*, McGraw-Hill Book Co., New York, 1950.
- Sonder, R. A., Ueber die Spannungsverteilung beanspruchten Kristallverbänden und deren Bedeutung für Gefügeregelung und Gesteinsmetamorphose, *Schweiz. mineral petrog. Mitt.*, **13**, 471-490, 1933.
- Sonder, R. A., Gefügeregelung, Druckschieferung und Druckverformung von Gesteinen, *Schweiz. mineral. petrog. Mitt.*, **28**, 362-379, 1948.

(Received August 10, 1961.)

Corrigendum

Enik Talwani, J. Lamar Worzel, and
Alice Ewing wish to point out some errors in
their paper 'Gravity Anomalies and Crustal
Structure across the Tonga Trench' which appeared
in the April 1961 issue of the Journal. In Figure 2
the air anomaly values at stations 66, 69, and

80 are 103, -168, and -173 mgal, respectively,
not -103, -108, and -17. In the legends of
Figures 4 and 7 and in the text on pages 1271-1275
where these figures are discussed, the station
numbers are 98, 99, 100, 101, 102, 103, 104, and
105, *not* 97, 98, 99, 100, 101, 102, 103, and 104.

Information for Contributors to the *Journal of Geophysical Research*

Manuscripts. Send manuscripts to J. A. Peoples, Jr., Department of Geology, University of Kansas, Lawrence, Kansas. Manuscripts, including proof copies of figures, should be submitted in triplicate to expedite review and publication. Manuscripts should be in English, typewritten on heavy paper on one side of page only, double spaced (including abstracts and references), with generous margins.

Ample space should be allowed for mathematical expressions, which should be typed or very plainly written by hand. Particular attention should be given to the legibility of subscripts and superscripts and to differentiation between capital and lower-case letters. Unusual symbols and cumbersome notation should be avoided. Fractional exponents should be written in preference to root signs, and the solidus (/) should be used for fractions wherever its use will save vertical space.

Authors are urged to have their papers critically reviewed by their associates for scientific validity, manner of presentation, and use of English before submitting them for publication.

Abstracts. An abstract must accompany each manuscript. It should be a concise but comprehensive condensation of the essential parts of the paper, suitable for separate publication and adequate for the preparation of general indexes to geophysical literature.

References and footnotes. References should be indicated in the text by the insertion in brackets of the author's name and the year of publication, thus: [Faust, 1958]. If the author's name is part of the text, only the year is bracketed. If there are two or more references citing different papers published in the same year by the same author, they may be distinguished by the letters *a*, *b*, *c* after the year.

At the end of the paper references should be listed alphabetically by the author's names in the form of the references given below. (See *List of Periodicals Abstracted by Chemical Abstracts* for abbreviations of titles of journals, or write titles in full.)

Footnotes to the text should be avoided; parenthetical sentences should preferably be added to the text. If footnotes must be given they should be numbered consecutively throughout the paper.

Tables and figures. Material suited to tabular form should be arranged as a table and may be typewritten on a separate sheet. Tables must be numbered according to their sequence in the text, and each table should have a title. Column headings should be short and self-explanatory; more complete explanation may be given in footnotes to the table.

Figures should be prepared with the column width of this Journal in mind (a scale of 2 to 4 times that of the published figure is usually adequate). They must be drawn in India ink on white paper or tracing cloth. Coordinate paper should be avoided; if used, however, it must be blue-lined, and coordinate lines that are to show must be inked.

Titles of figures must be typewritten consecutively on a separate sheet; they are *not* to be lettered on the figure. Necessary lettering within the figure must be executed to meet competent drafting standards and should be large enough to remain legible after reduction. The ideal letter size in the reduced figure is 1/16 inch, but 1 mm is acceptable. Unnecessary blank space within the figure should be avoided.

Photographs are acceptable only if they have good intensity and contrast. They should be unmounted glossy prints.

Figure number and author's name should be written lightly in pencil on each figure. 'Top' of each figure should be indicated.

Acknowledgments. Acknowledgments should be made only for significant contributions by the author's professional associates. A brief closing statement will usually suffice.

REFERENCES

- American Chemical Society, *List of Periodicals Abstracted by Chemical Abstracts*, Chemical Abstract Service, Ohio State Univ., Columbus, 314 pp., 1956.
- American Institute of Physics, *Style Manual*, 2nd ed., American Institute of Physics, New York, 42 pp., 1959.
- Faust, L. Y., The preparation of a paper, *Geophysics*, 23, 944-952, 1958.
- Landes, K. K., A scrutiny of the abstract, *Geophysics*, 17, 645, 1952. Reprinted in *Geophysics*, 23, 942-943, 1958.
- Skillen, M. E., R. M. Gay, and others, *Words into Type*, Appleton-Century-Crofts, New York, 585 pp., 1948.
- U. S. Geological Survey, *Suggestions to Authors of the Reports of the United States Geological Survey*, 5th ed. U. S. Govt. Printing Office, Washington, 255 pp., 1958.
- William Byrd Press, *Mathematics in Type*, William Byrd Press, Richmond, 58 pp., 1954.



CORPORATION AND SUPPORTING MEMBERSHIPS

The American Geophysical Union is a non-profit scientific organization established by National Research Council. Its Council is the United States National Committee of International Union of Geodesy and Geophysics; official adherence by the United States is through the National Academy of Sciences-National Research Council.

Extracts from the Statutes:

Article 3. Membership—The membership of the American Geophysical Union shall be as follows:

- (e) *Corporation Members*—Corporations and other organizations interested in geophysics elected by the Council of the Union. The designated representative of each such organization shall enjoy the privileges of a Member.
- (g) *Supporting Members*—Corporations, other organizations, and individuals interested in geophysics and desirous of supporting the Union may become members under the following classifications upon election by the Council.

(Continued on next page)

Cut along this line

American Geophysical Union

PROPOSAL FOR _____ MEMBERSHIP

to the Council, American Geophysical Union

1515 Massachusetts Avenue, N.W., Washington 5, D. C.

gentlemen:

As an indication of our interest in the aims and activities of the American Geophysical Union, and to assist in maintaining and extending its program of publication and other work in the development of the geophysical sciences, the undersigned applies for _____ Membership in the AGU and, until further notice, agrees to pay annual dues at the rate established for this classification of membership, in accordance with the information set forth above and on the following page.

Company or Organization _____

(Signature)

Title _____

(over)

(Continued from previous page)

of the Union . . . : Contributing Members, . . . Sustaining Members, . . . Benefactors. . . . Each Supporting Member and the designated representative . . . shall enjoy the privileges of a Member.

Corporation Members shall pay dues of not less than \$100 for each calendar year. Dues for Supporting Members shall be as follows:

Contributing Members	\$500
Sustaining Members	\$1000
Benefactors	\$5000

Lists of Corporation Members and Supporting Members will be published in each issue of the *Transactions*, and will be included in the Membership Directory as distinct units.

By Laws provide that one copy of each issue of the *Transactions*, *Journal of Geophysical Research*, any published *List of Members and Officers*, and any other publication which may be approved for free distribution to the membership shall be sent to each Corporation and Supporting Member. Each organization in good standing may purchase any available publication of the Union at the established member discount.

AMERICAN GEOPHYSICAL UNION

1515 Massachusetts Ave., N.W.
Washington 5, D. C.

Cut along this line

Address _____

City _____ State _____

General fields of activity _____

The following person is designated as our representative in this membership _____

_____ Title _____

Number of units of membership desired (this will be taken as one unless otherwise indicated) _____

Place _____

Date _____

AMERICAN
GEOPHYSICAL
UNION

UNSELFISH
COOPERATION
IN RESEARCH

AMERICAN GEOPHYSICAL UNION

1515 Massachusetts Avenue, N.W., Washington 5, D. C.

Established by the National Research Council in 1919 for the development of the science of geophysics through scientific publication and the advancement of professional ideals.

QUALIFICATIONS FOR MEMBERSHIP

Membership of the AGU shall consist of Members, Associate Members, Student Members, Corporation Members, and Supporting Members.

Those eligible as candidates for election to the grade of MEMBER shall be:

MEMBER (a) Persons who have made an active contribution to geophysical research through observation, publication, teaching, or administration. Definite evidence should be presented to the Membership Committee. "Publication" may include books, articles, unpublished manuscripts, inventions, or development of geophysical instruments.

(b) Persons who have made active practical application of geophysical research. It should be shown that the nominee's work has not been purely routine, but that it has tended to create new knowledge of, or to broaden or strengthen the application of, geophysical research. In general, the minimum qualifications for membership will be not less than three years of professional experience in some phase of geophysics.

(Continued on next page)

Cut along this line

APPLICATION FOR MEMBERSHIP

Please refer to qualifications on reverse side and designate below type of membership desired:

Member (\$10) ☐

Associate (\$10) ☐

Student (\$4.50) ☐
(1961)

Application forms for Corporation and Supporting Membership are available upon request.

Surname

First Name

Middle Name

Preferred mailing address for publications

Permanent address

Place Month Day Year of Birth

4.

Country of citizenship/naturalization

Nature of work and title and/or military rank; name and address of organization with which you are associated.

Check section or sections with which affiliation is desired.

☐ Geodesy

☐ Seismology

☐ Meteorology

☐ Geomagnetism and Aeronomy

☐ Oceanography

☐ Volcanology, Geochemistry, and Petrology

☐ Hydrology

☐ Tectonophysics

EXPERIENCE (List below, use added sheets as necessary)

Dates: From To Name and address of organization Title, duties, nature of work

EDUCATION (List Below, use added sheets as necessary)

Dates: From To School Address Major Subject Degree, if any; year

(over)

(Continued from previous page)

Those eligible as candidates for election to the grade of ASSOCIATE MEMBER shall be:

ASSOCIATE MEMBER Persons who have an active interest in physical processes of the Earth or technical assistance in the application of geophysics. In general, the minimum qualification for associate membership will be acceptable training or experience in some field of geophysics or allied science.

CORPORATION AND SUPPORTING MEMBER Corporations and other interested organizations shall be eligible as candidates for election to CORPORATION or SUPPORTING MEMBERSHIP. They shall have the privilege of designating a representative who has the rights and privileges of Members (use special form).

STUDENT MEMBER Those eligible as candidates for election to the grade of STUDENT MEMBER shall be persons who are graduate or undergraduate students in residence at least half-time and who are specializing in the geophysical sciences. Teaching or research assistants enrolled in more than half of a full-time academic program may also be eligible for Student Membership. Student Members shall have all the privileges of Members except that they shall not vote or hold office.



Cut along this line

*9. References: Please list below names and addresses of two or three references; include members of the AGU or others who know you well.

*10. Titles of technical contributions or publications, particularly those in the geophysical sciences, and where published.

*11. Brief statement of any special interests or qualifications in the geophysical sciences.

Date _____

Written Signature

12. (STUDENT MEMBERS ONLY) The person whose signature appears above is known to me and is a student majoring in _____ (subject) at _____

(Name of college or university) expected to graduate in _____ (year) with the degree of _____

☐ He is a full-time student, or ☐ a teaching or research assistant enrolled in more than half of a full-time academic program.

(Signature of faculty sponsor)

☐ Check here if faculty sponsor is a member of AGU and willing to act as a regular sponsor for associate membership as well.

(Typed or printed name of sponsor)

(Title)

* Applicants for student membership may omit Questions 9, 10, and 11, but must fill in Question 12. Please return form with check or money order payable to American Geophysical Union, 1515 Massachusetts Ave., N.W., Washington 5, D. C.

Contents

(Continued from back cover)

	PAGE
Seismetering Ocean-Bottom Seismograph.....	<i>John Ewing and Maurice Ewing</i> 3863
Evaluation of a Signal-Summing Technique for Improving the Signal-to-Noise Ratios for Seismic Events.....	<i>L. H. Koopmans</i> 3879
Oscillations of the Moon.....	<i>H. Takeuchi, M. Saito, and N. Kobayashi</i> 3895
Seismic Motion Arrivals along a Thin Elastic Plate Surrounded by a Fluid Medium	<i>J. H. Rosenbaum</i> 3899
Geophysical Requirements in Terrestrial Expansion..	<i>Melvin A. Cook and A. J. Eardley</i> 3907
Stability Measurements of Rock Salt..	<i>Ernest F. Gloyna and Tom D. Reynolds</i> 3913
System $\text{NaAlSi}_3\text{O}_8\text{-H}_2\text{O-Argon}$: Total Pressure and Water Pressure in Metamorphism.....	<i>H. J. Greenwood</i> 3923
Magnetic and Solar Data.....	<i>J. Virginia Lincoln</i> 3947
Letters to the Editor:	
The Primary Cosmic-Ray Electron Flux during a Forbush-Type Decrease	<i>Peter Meyer and Rochus Vogt</i> 3950
Geomagnetic Activity and the Reception of Whistlers in Polar Regions	<i>G. McK. Allcock and M. F. Rodgers</i> 3953
Discussion of Paper by John R. Spreiter and Benjamin R. Briggs on 'Theory of Electrostatic Fields in the Ionosphere at Polar and Middle Geomagnetic Latitudes'.....	<i>D. T. Farley, Jr.</i> 3956
Increase of Ionization Associated with Geomagnetic Sudden Commencements	<i>S. Matsushita</i> 3958
Two Methods of Detecting Ionospheric Disturbances	<i>G. A. M. King and M. D. Lawden</i> 3962
Note on 'Vertically Traveling Shock Waves in the Ionosphere'	<i>Fred B. Daniels and Arthur K. Harris</i> 3964
Discussion of Incoherent Backscatter Power Measurements at 440 Mc/s	<i>V. C. Pineo and H. W. Briscoe</i> 3965
Artificial Initiation of Lightning Discharges	<i>M. Brook, G. Armstrong, R. P. H. Winder, B. Vonnegut, and C. B. Moore</i> 3967
Interplanetary Dust Particles of Micron-Size Probably Associated with the Leonid Meteor Stream	<i>W. M. Alexander, C. W. McCracken, and H. E. LaGow</i> 3970
Some Micrometeorological Measurements in Ocean Fog	<i>Francis M. Wiener, Jay H. Ball, and Creighton M. Gogos</i> 3974
Reply to Dr. Harold C. Urey's Criticism of the Paper by Brian Mason, 'The Origin of Meteorites'.....	<i>Brian Mason</i> 3979
A Note on the Thermodynamic Theory of Nonhydrostatically Stressed Solids	<i>Mineo Kumazawa</i> 3981
Author's Reply to Discussions of the Paper 'The Thermodynamic Theory of Nonhydrostatically Stressed Solids'.....	<i>W. Barclay Kamb</i> 3985
Appendix.....	<i>Manik Talwani, J. Lamar Worzel, and Maurice Ewing</i> 3989

Contents

Formation of the Geomagnetic Storm Main-Phase Ring Current	
	<i>A. J. Dessler, W. B. Hanson, and E. N. Parker</i>
Spectrum of Hydromagnetic Waves in the Exosphere. . . .	<i>Gordon J. F. MacDonald</i>
Spatial Distribution of Electrons from Neutron Decay in the Outer Radiation Belt	
	<i>W. N. Hess and J. Killeen</i>
Coordinates for Mapping the Distribution of Magnetically Trapped Particles	
	<i>Carl E. McIlwain</i>
A Theoretical Model of Temperature Variations at the Surface of an Orbiting Satellite.	<i>Heinz H. Lettau</i>
Propagation Characteristics of Whistlers Trapped in Field-Aligned Columns of Enhanced Ionization.	<i>R. L. Smith</i>
Properties of the Outer Ionosphere Deduced from Nose Whistlers. . . .	<i>R. L. Smith</i>
Concerning Radiosondes, Lag Constants, and Radio Refractive Index Profiles	
	<i>B. R. Bean and E. J. Dutton</i>
The Solar Semidiurnal Atmospheric Oscillation.	<i>K. A. Small and S. T. Butler</i>
Traveling Pressure Waves Associated with Geomagnetic Activity	
	<i>Peter Chrzanowski, Gary Greene, K. T. Lemmon, and J. M. Young</i>
Measurements of Current Density in the Fair Weather Atmosphere	
	<i>J. H. Kraakevik</i>
Atmospheric Emission and Opacity at Millimeter Wavelengths Due to Oxygen	
	<i>M. L. Meeks</i>
Chemical Composition of Rain Water in Kampala, Uganda, and Its Relation to Meteorological and Topographical Conditions.	<i>Simon Visser</i>
Tritium Geophysics.	<i>W. F. Libby</i>
Stability of Ice-Age Ice Sheets.	<i>J. Weertman</i>
Depth to Sources of Magnetic Anomalies	
	<i>Leroy R. Alldredge and Gerald D. Van Voorhis</i>
Effects of Moderate Stresses on Directions of Thermoremanent Magnetization	
	<i>John W. Kern</i>
The Effect of Stress on the Susceptibility and Magnetization of a Partially Magnetized Multidomain System.	<i>John W. Kern</i>
Stress Stability of Remanent Magnetization.	<i>John W. Kern</i>
Observations of Internal Waves near Hudson Canyon.	<i>Roy D. Gaul</i>
Periodicity of Earthquakes and the Invariance of the Gravitational Constant	
	<i>W. J. Morgan, J. O. Stoner, and R. H. Dicke</i>
Some Characteristics of Surface Gravity Waves in the Sea Produced by Nuclear Explosions.	<i>Wm. G. Van Dorn</i>

(Continued inside back cover)